Table 1

SEQ ID NO:5 through SEQ ID NO:39

SEQ ID NO:#	Sequence
SEQ ID NO:5	ATCTAACAAT TGGTTCAAGT CGCTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAGC GTTAGGT
SEQ ID NO:6	ACCTAACATG GCGCTCAACC GCGCTCCCTT CGGTCGCTGG ACGCTGCGCG ATAAAGCCGC GCGGCCGCCC TTAGCTCTAC CGTTAGC
SEQ ID NO:7	GTCCTAACAAT TGGCTCAAGT CGTTCGCTTC GCCGCTCACT TCGTTCGCTG GACTCGCAAA AGCTGCGCTT TTGCTCGCCC GTTAGCTTAA TCLTTAC
SEQ ID NO:8	ACCTAACAAT GCGCTCAACT GCCGCTCACT TCGTTCGCTG GACAGTCAAA AGCTGCGCTT TTGCTCGCCC GTTAGCTTAA TCGTTAC
SEQ ID NO:9	ACCTAACAAT GCGCTCAACT CGCTCACTTC GTTCGCTGGG ACCGGCGAAG CCGGCCCCTT AGCTTAATCG TTAGGT
SEQ ID NO:10	GCCTAACAAG GCGCTCAAAG CGCTCACTTC GTTCGCTGGG ACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGT
SEQ ID NO:11	ACCTAACAAC TGGTTCAAGT CGTTCGCTTC GCTCACTCGG GACCGGCGAA GCCGGCCCCT TAGCCAAACG TTAGGT
SEQ ID NO:12	GCCTAACAAT TGGCTCAAGT CGTTCGCTTC GCCGCTCACT TCGTTCGCTG GACAGTCAAA AGCTGCGCTT TTGCCTGCCC GTTAGCTTAA TCGTTAC
SEQ ID NO:13	ACCTAACAAT GCGCTCAACT GCCGCTCACT TCGTTCGCTG GACAGTCAAA AGCCGGCCCT TAACCAAACG TTAGGC
SEQ ID NO:14	GCCTAACAAG TGGTTCAAAC CGTTCGCTTC ATCGCTCACT TCGTTCGCTG GACTCGCAAA AGCCGCGCTT TTGCTCGCCC GTTAGCTTAA TCGTTAT
SEQ ID NO:15	CTCTAACAAT GCGCTCAACT ATCGCTCACT TCGTTCGCTG GACTCGCAAA AGCCGCGCTT TAGCTTAATC GTTAAAT
SEQ ID NO:16	TTATAACAAT GCGCTCAACT CGTTCGCTAC TCGTTCGCTG GACAGCCAAA AGCTGCGCTT TTGTCTGCCC GTTAGCTTAA TCGTTAC
SEQ ID NO:17	ATTTAACAAT GCGCTCAACT GTCGCTCACT TCGTTCGCTG GACAGCCAAA CCTTAACCAA CGTTAGAG
SEQ ID NO:18	CCCTAACAAA TGGTTCAAGT CCGTTCGCTT CGCTCACTCG GGACCGGCTA AAGCCGGCCC CTTAACCAAA CGTTAGGC
SEQ ID NO:19	CTCTAACAAA TGGTTCAAGT CGCTCGCTTC GCTCACTGG ACTCGCAAAA GCTACGCTTT TGCTCGCCGG TTAGCTCAAA CGTTAGC
SEQ ID NO:20	GCCTAACAAC TCACTCAACC TCGTTCGCTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGG
SEQ ID NO:21	GCCCAACAAA TGGCTCAAGT CGCTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGC
SEQ ID NO:22	CCCTAACTAG TGGTTCAAGC CGTTCGCTTC GCTCACTCGG GACCGGCTAA GGACCGGCCC TTAACCAAA CGTTAGGT
SEQ ID NO:23	GCCTAACAAT TGGTTCAAGT CGTTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAA CGTTAGGC
SEQ ID NO:24	ACCTAACAAT TGGTTCAAGT CACTCGCTTC GCTCGTTCGG GACCGGCATA GCCGGCCCCT TAACCAAACG TTAGGT
SEQ ID NO:25	GCCTAACAAC TGGTTCAAGT CGTTCGCTTC ACTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGT
SEQ ID NO:26	GCCTAACAAC GCGTTCAAGT CGTTCGCTTC CGCTCACTCC GGACGCCCGC AAGCTACGCT CGGGGTCGCC CCTTAACCTG TCCGTTA
SEQ ID NO:27	GCCTAACAAA CGGTTCAAGT TCGTTCGCTT CGCTCACTTC GACGTCAAA AGCTGCGCTT TTGGCCGCCC GTTAGCTTAA CCGTTAT
SEQ ID NO:28	ACCTAACAAT ACGCTCAACT ATCGCTCACT TCGTTCGCTG GACAGCCAAA AGCTACGCTT TTGCCTGCCC GTTAGCTTAA TCGTTAC
SEQ ID NO:29	ACATAACAAT GCGCTCAACT GCCGCTCACT TCGTTCGCTG GACAGCCAAA AGCTGCGCTT TTGTCTGCCC GTTAGCTTAA TCGTTAC
SEQ ID NO:30	GCCTAACAAG TCGCTCAACT CCGTTCGCTG GACGTCCAAA AGCTGCGCTT TTGGCCGCCC GTTAGCTTAA TCGTTAC
SEQ ID NO:31	GCCTAACAAT GCGCTCAACT ATCGCTCACT CCGTTCGCTG GACGTCCAAA GCCGGCCCCT TAACCAAACG TTGGGC
SEQ ID NO:32	GCCTAACAAC TGGTTCAAGC CACTCGCTTC GCTCGCTCGG GACCGGCTAC CGGCGCCCCT TAACCAAA CGTTAGGG
SEQ ID NO:33	GCCCAACAAA CGGTTCAAGA CCGGCTCGCCT TGCTCGCTCG GGACCGGCTAA AAACCGGCCC CTTAACCAA TAGGC
SEQ ID NO:34	GCCTAACAAT TGGTTCAAGT CGCTCGCTTC GCTCGGCTGGG ACCGGCGAAG CCGGCCCCTT AACCAAACGT TTAGGT
SEQ ID NO:35	GCCTAACAAT GCGCTCAAAG CGCTCACTTC GTTCGCTGGG ACCGGCCAAA AGCTGCGCTT TTGCCTGCCC GTTAGCTTAA CCGTTAT
SEQ ID NO:36	ACCTAACAAT GCGCTCAACT TCGTTCGCTG GACGGGGCC CTCAAACGTT ATAA
SEQ ID NO:37	CGCTAACAAT TCGCTGCAGG CGCGACGGCC GCGGCCTTGAG AAGCCGGCTA GGACCGGCGA CTTAACCAAA CGTTAGGC
SEQ ID NO:38	TTATAACAAT TGGTTCAAGT CGTTCGCTTC GCTCACTGCG GGACCGGCTA CGCCGGCCCC TTAGCTTAAT CGTTAGGC
SEQ ID NO:39	GCCTAACAAA TGGTTCAAGT CGCTCATTCGG GCTCACTCGG GACCGGCTAA GCGCGGCCCC

Table 1 Continued
SEQ ID NO:40 through SEQ ID NO:75

| SEQ ID NO	Sequence
SEQ ID NO:40	GCCTAACAAT AGGTTCAAGT CGCTCGCTTC GCTCACTTGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGT
SEQ ID NO:41	ACCTAACTAG TGGTTCAAGC CGCTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGC
SEQ ID NO:42	ATATAACAAT TGGTTCAAAC CGTTCGCTGC GCTCACTGGG ACGGGCTAAA GCCCGCCCCT TAACCAAACG TTATGC
SEQ ID NO:43	GCATAACAAT TGGCTCAAGC CGCTCGCTCC GCTCACTCGG ACGTCCGTAA GCTACGCTTC CGGCCGCCCC TTAGCCAAAC GTTAGGC
SEQ ID NO:44	CCCTAACAAA TGGTTCAAAG CCGTTCGCTT CGCTCACTCG GGACCGGCTA AAGCCGGCCC CTTAACCAAA CGTTAGGC
SEQ ID NO:45	GCCTAACAAT GCGCTCAACT GCCGCTCACT TCGTTCGCTG GGACAGTCAAA AGTCGCGCTT TTGCCTGCCC GTTAGCTTAA TCGTTAG
SEQ ID NO:46	CTCTAACAAT GCGCTCAACT CGCTCACTAC GTTCGCTGGG ACCGGCTAAA GCCGGCCCCT TAGCTTAATC GTTAGAG
SEQ ID NO:47	CTCTAACAAT TGGTTCAAGT CGTTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC CTTAACCAAA CGTTAGGG
SEQ ID NO:48	CCCTAACAAA TGGTTCAAGT CACTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGC
SEQ ID NO:49	CTCTAACAAC TGGTTCAAGT CGTTCGCTTC GCTCACTCGG GGACCGGCTA TTGGCCGCCC CTTAGCCAAA CGTTAGGG
SEQ ID NO:50	ATCTAACAAT TGGCTCAAGC CGTTCGCTTC GCTCACTCGG GACGTCCAAT AGCTGCGCTA CCTTAACCAA ACGTTAGGC
SEQ ID NO:51	GCCTAACAAC TGGTTCAAGT CGTTCGCTTC GCTCACTGCC GGACCGGCTA AAGCCGGCCC CCTTAACCAA ACGTTAGGC
SEQ ID NO:52	CCCTAACAAA TGGTTCAAGT CGTTCGCTT GGACACCAAA CTGTACGCAG TTTGGTTCCC CCGGTGCCG GTTACTTTCA ACGTTAG
SEQ ID NO:53	GCCTAACACT GCAGTCAACC GCTCACTTC GTTCGCTGGG ACCGGCGAAG CCGGCCCCTT AGCTTAATCG TTAGCC
SEQ ID NO:54	GCCTAACAAT GCGCTCAAAG GCGCTCACTTC GTTCGCTGGG ACGGCTAA AGCCGGCTAA TTAACCAAAC GTTAGCC
SEQ ID NO:55	TTCTAACTAC TGGTTCAAGT GTCGCTCACT CGGACGCAAA AGTCACGCTT TTGTCTGCCC GTTAGCTTAA TCGTTAG
SEQ ID NO:56	GGCTAACAAT GCGCTCAACT GTCGCTCACT GCTCGCTCGG GACAGCCAAA AGTACGCTT CGCGCGCCCT TAACCAAACG TTGGGC
SEQ ID NO:57	GCCTAACAAC TGGTTCAAGC CACTCACTTC GCTCGCTCGG GACGTCCGAA AGTACGCTT TTGGCCGCCC GTTAGCTTAA TCGTTA
SEQ ID NO:58	GCCCAACAAT GCGCTCAACT GCCGCTCACT TCGTTCGCTT GGACCGGCTA AAGCCGGCCC CTTAACCAAA CGTTAGGC
SEQ ID NO:59	GCATAACAAT TGATTCAAGT CGTTCGCTTC GCTCACTGCG GCTCGTTCGG GACCGGCGTT CGCGGCCCCT TAACCAAACG TTATGC
SEQ ID NO:60	GCCTAACTAC TGGTTCAAGT CACTCGCTTC GCTCACTCGG GACCGGCGAA GCCGGCCCCT TAGCCAAACG TTATGC
SEQ ID NO:61	ATCTAACAAT TGGCTCAAGT CGCTCGCTTC GCTCACTCGG ACGTCCGTAA GCTACGCTT CGGCCGCCCC TTAACCAAAC GTTAGGC
SEQ ID NO:62	GCATAACAAA TGGTTCAAGC CGTTCGCTTC GCTCACTCGG ACGTCCGTAA GCTACGCTTC CGGCCGCCCC TTAACCAAAC GTTAGGC
SEQ ID NO:63	GCCTAACAAA TGGTTCAAGT GGACGCAAAC CCCGGCTCGC GGTCTTTGCG CCGGCTTATCT CAAGCGTTAG AT
SEQ ID NO:64	GCCTAACTAT TCAGTCAAGC CGCTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGAG
SEQ ID NO:65	ATCTAACATG TGGTTCAAGC CGCTCGCTTC GCTCGCTCGG GATCGGCGAA GCCGGCACCT TAACCAAACG TTAACT
SEQ ID NO:66	CTCTAACAAT TGGTTAAAAC CGTTCGCTTC GCTCACTCGG GACCGGCGTG AAGCCGGCCC CTTAACCAAA CGTTAGGC
SEQ ID NO:67	CTCTAACAAT TGGTTCAGAT CGTTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGGC
SEQ ID NO:68	GCCTAACTAC TGGTTCAAGT CGCTCACTTC GCTCACTCGG ATCGGCTGGG ATCGGCCTAA AGCCGGCCCC TAGCTTAATC GTTAGCA
SEQ ID NO:69	GTCTAACAAT GCGCTCAAAG CGCTCGCTTC GTTCGCTGGG GATAGTCAAA AGTCGCGCTT TTGTCTGTCC GTTAGCTTAA TCGTTAC
SEQ ID NO:70	TGCTAACAAT GCGCTTAACT GTCGCTCACT TCGTTCGCTG GATAGTCAAA CCGGCCCTTA ACCAAGCGTT AGAT
SEQ ID NO:71	GCCTAACAAC TGGTTCAAGT CGCTCGCTTC GCTCACTCGG ACCGGCATAG CCGGCCTAAA GCCGGCCCCT TAACCAAACG TTAGGT
SEQ ID NO:72	ATCTAACAAT TGGTTAAAAC CGCTCGCTTC GCTCACTCGG GATCGGCTAA ATTCGGCCGAA TTCGGCCCCT TAGCAAACG TTAACT
SEQ ID NO:73	GTTTAACAAC TGGTTCAAGC CGCTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAAACCAAG CGTTATGC
SEQ ID NO:74	ATCTAACAAT TGGTTCAAGT CGCTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGAT
SEQ ID NO:75	GTATAACAAT TGGTTCAAGT CACTCGCTTC GCTCGCTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGAT

Table 1 Continued
SEQ ID NO:76 through SEQ ID NO:78

SEQ ID NO:76	GCATAACAAG TCGCTCAAAT CGCTCACTTC GTTTCGCTGG GACGGGCTAA AGCCCGCCCC TTAGCTTATC GTTAGGC
SEQ ID NO:77	GCCTAACAAA TGGTTCAAGC CGTTCGCTTC GCTCACTCGG GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGAC
SEQ ID NO:78	CTCTAACAAA TGGTTCAAGC CGTTCGCTTC GCTCACTCGG GACCGGCTAA AGCCGGCCCC TTAACCAAAC GTTAGAG

09/701626
PCT/US99/13295

Table 2

SEQ ID NO:79 through SEQ ID NO:91

	Name in text	strand	Sequence
SEQ ID NO:79	1	coding	TCGCTTCGCT CACTGCGGGA CCGGCTAAAG CCGGCCCCTT AACCAAACGT TA
SEQ ID NO:80	2	coding	TAACAATTGG TTCAAGTCGT TCGCTTCGCT CACTGCGGGA CCG
SEQ ID NO:81	3	coding	TAACTATTCA GTCAAGCGGA CGCAAACCCC GCTGCGCGGT CTT
SEQ ID NO:82	4	coding	TAACAATGCG CTCAACTGCG CTCACTTCGT TCGCTGGACA GCC
SEQ ID NO:83	5	coding	TAACAAGTCG CTCAACTGCC GCTCACTCGT TCGCTGGACA GCC
SEQ ID NO:84	6	noncoding	GCCCCTTAAC CAAACGTTA
SEQ ID NO:85	7	noncoding	CCGAGTGAGC GAAGCGAGCG
SEQ ID NO:86	8	noncoding	AAACTCGAGG GTCCCGAGTG AGCGAAGCGA GCG
SEQ ID NO:87	9	noncoding	AAACTCGAGG GTCCCGAGCG AGCGAAGCGA GCG
SEQ ID NO:88	10	noncoding	AAACTCGAGG CTGTCCAGCG AGCGAAGCGA GCG
SEQ ID NO:89	11	noncoding	AAACTCGAGA CCGCGCAGCG GGGTTTGCGT CCG
SEQ ID NO:90	12	coding	TGCTCTAGAC GGCCCCTTAA CCAAACGTTA G
SEQ ID NO:91	13	coding	TGCTCTAGAC GGCCGCCCGT TAGCTTAATC GTTAG

Fig 1 Integron structure
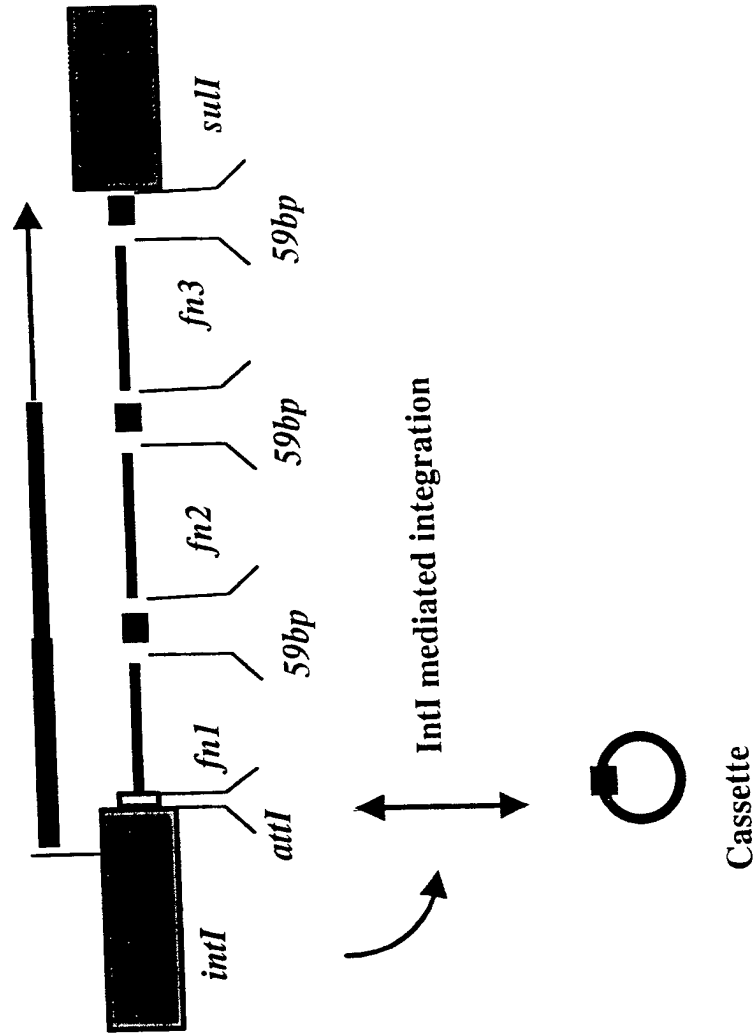
after Hall and Collis 1995

Fig. 2

Vibrio cholerae superintegron
fragment carried on pPM147

BamHI

BamHI
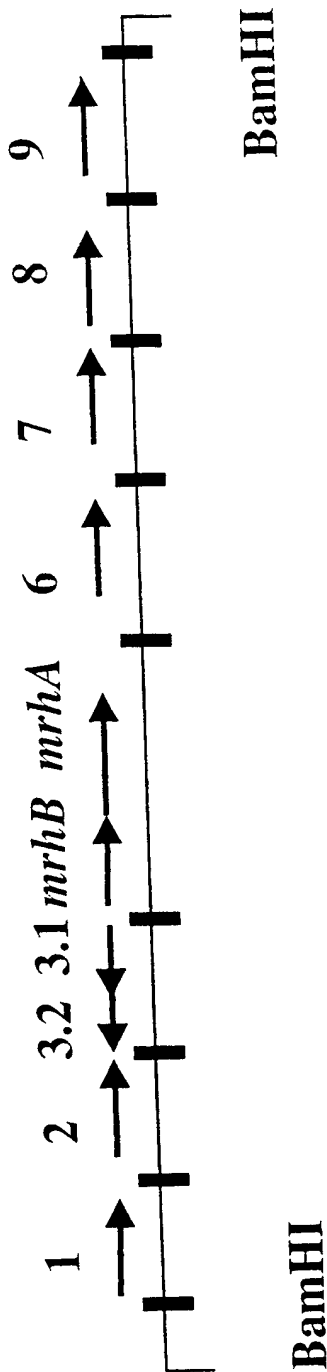
| VCR repeat

→ Open reading frame (ORF)

ORF2 is similar to Vlp()

ORF3.1 is similar to a plasmid-encoded protein (gi|516610)

ORF3.2 is similar to RelE (gi|42701) and plasmid-encoded proteins

Figure 3A-1
SEQ ID NO:1

```
ATCGATCAGC CAGACTTTTC GCACACGGGC GGACCTTGGG CGAGTCAGCG
CTATGGTTGG CCGCTGTGGG TTGTCAGTGC CCGTACGCGC AATCTGTTTC
TTTCGCAGGG CATGTCCGGC TGGGCGTTCC GGCCCGTTCT GGTCACCGAC
TCGGCTCTCT ATGAGCGCTA TCTCGCTCTA AGTCAGGAAC TTTGCGCACT
GCTTCGTGAT GCACCGCAGA GCAAGCTCGA AGACCGTGAT TGGTAAGCGG
GGGCTATTCG ATCAGTCTCG GAGCGACCAA ACTCCAGAAA CGACAAGGCC
CTGAAAAAAA AGCAGGGCTT CGTCTTTGCG GGCGAATGGA ATCGGACCTC
TTTCCGCCTC TGCATGTAAC TGGTCTTTGT TTGCCAAATC TGCCTATCTC
ATGCCGGCCA TGTTGGCCAG TGCCTGCATC ATTTGGCCTT TGGTTTCGAC
ACTTTTTCGA CAGCCCTGCT AGACATCCCT CCCTCTGCCC TCGTAACTTC
TGTTCCGATG GTGTCGCTTG GCACTATGGT CTTGTCGAGT GTCGCTTTTC
ATCCAGCCTA ATGCCGCGAT TGCCTCGCTG AGCTGTAGCT GAATCAAGGA
CTTAGCGGAC GACAAGGAAT GTTATGCGAA ACATGTGGCG GAATAAATTA
CGCCGCATGT TTCGTCTACT TATAGTTAGG CTACATATGA GAATCAGCGC
AGACCAGCTT GCTCAAGAAT CACTGACTGA GTTCGGCGTG CTGGCGGCTA
AGCTTCTGGC AACGCGAGAG CTTAGCCAGT TGTCCGAGAA GTTTGGGTAT
GCACTGGCCT TCGGAAGGGA ACCGGCGGCT GCCATAGCTG AGGACCTTGC
TAGGTGCTTG TGCGGACAAA ATGCTTCGCC GGCATCTGAA TACCCCAAAA
TCACCGTTAA GTATTTCAAG GAAAACGAAA GTAGTCTGTT GGCACTCGTA
GAGTGTTATG TACAAATGAC CGCAAGCGCA AACATTCTTT TAGAGCTGGT
TGCCGCACGA AATGGAGAGG CAATAAATCT GTATCTAGAA GGCTTGAGTG
TTGTAGCCTA ACAATGCGCT CAAAGCGCTC ACTTCGTTCG CTGGGACCGG
CGAAGCCGGC CCCTTAGCTT AATCGTTAGA AACCATCATG GATAACTGGT
ACAACACCAT CGAATACCAA ACCCATGTAG CCGAAAAACT AGAGGCACTT
GGAGAAACAA AGTACGACCG CGAGGCTTAT GAATTCGCGC TAGAGGCATA
CCAGTATGCG CCTGAATATC ATGAAAATAT TCCCACGCCG CCTCTCAATC
TTGGGCTCGC GTACCATGTA AGCGCCTTCA ACTTTGCACA CTGCTATGTA
CTTCACGCTA AAGAAGTGTT TGAAGCTCCA AAAGACACAC TGAGCTCCTG
GGGCGTATTT TCCTCAACGG ACATTGGTGA AATTGTTTAT GGTTTAGTCC
GTATTGGCTT GCTGGACCAA GGCCCCGAAG ACAAAAAAGA GCAGTTTGAA
GGGTTGTTTT TAATCACCGA CGTGCTGTGA TGTCTTCTAA CTACTGGTTC
AAGTCGTTCG CTTCGCTCAC TCGGGACCGG CTAAAGCCGG CCCCTTAACC
AAACGTTAGC CACCTCACGA AGATTTGGAG CCCGCGTGAA CAAAGTCGAT
ACAAACAAAA TTAAAACGGA TTTTTCGGCA CGAATTGATG AAAAAAGAGC
GTGGTTTGAT CGTATGGCTA CGCTTATAAG CGGGACAAAC ACCGAGTTAA
CCGACCTTAA TTTTCTTTGC GAGAACTATA TAACATCAAT ATACGTAGAG
CTCGAATGCT TAATATCAGA TTTATTTCAT GGCTACATAA ATAACAACAA
CAAGACCTAC ATGGCGCACA TTCAATCCATA CAACATTCGC AGGTCCAGAG
ACAAGTACTC TGCATGGCAC GCCACCGCTC CTTGATCCAA CAAGCTGGAA
CATATTAATT CAGCACAGCT CAGCACGTCT TAAGTGTGGG
CATCACATTT AAAGACGTTT CCGCAATGAA AGTACGAGCA AAGGAATACC
TTTCCTCAGT ACACGAAAAA AGATTTTCAG GTATATCTGC ATCCGATGGA
GCTCTTATTG ATGCCGCACA TGCAATCAGA AATTGCATTG CACACAACAG
CGAAAGCTCC AGAAAGGTTA TGAACACCAA AATTAAAAGC TTAATTACAG
GCCCAGCTTG CTCAAATGTC GGCCTTGAAC TCACCACAAA TAGTGTGACC
AAAATAGGAA AGTATCTCCG TGCAAATGCT CAGCAAAGCA TGCGAGTGCT
GATTTACTCA GATCGAATAA AATCTATCGG CCTAAGCTTA TAAGTGTGGG
CTAACAATGC GCTCAACTGT CGCTCACTTC GTTCGCTGGA CAGCCAAAAG
CTACGCTTTT GTCTGCCCGT TAGCTTAATC GTTAGGAGGC TCTGCATGAC
TCGTGCAACA GACAGGTTCG AAGAGCTTCT GCAATCACAT GAGTTCTCAG
GGCATATTAT TCGTTGGGTT GCGATATTCG AAGGCCGTCT TGACGGTGTG
TTATCAGTTC ATTTTTCTGG ACTTGAAAGC ACCTATGAAT TCTACGAACT
```

```
CATACTTTCC AGGTTGTCTT TCTACGAAAA AATTGAAATC CTGAGAAAAA
TTGATTTTGG TAACAGTCTC AAATCCCAAG AAAATACAGC GCTGCACCTA
GACAAACTGA GGCGATTGCG TAACGCATTG GCGCATGCAG CACACATGCC
ACCTGATGAA ATCATGAAGT TGTGCTCTGA TAAGTGGATA GAGTCCTTTG
TGCTCGGATA TCCAAAGTCC ATTGGCAAAG AGAAAAATGC ACTTGAAAAT
CGGCTATCAC TTCTGTGGAA TTACTGCCAC AGGAGGCATG TAGCAAAAAT
TAAGCAGCTT GCACACGAAC TCAAAAATAC AGAGCAAGCC AACTAATAGA
GTCCAGTTAT ACAGGTCCGT AAATGAGCCG CCTAACAACT GGTTCAAGCC
ACTCACTTCG CTCGCTCGGG ACCGCGTTCC GCGGCCCCTT AACCAAACGT
TGGGCACCCA TAGAAAAATC CTAATGAGAA AACTATTCAT ACCACTAATT
TTCGCCCTGC TATCGGAGAG CTTGATGGCA TCTGAAGCGT ATAAGGACCT
TGAAACACAA GTAACTGAAA AAGCCAGCCT AGCAGTTGCC CAAATGAATG
ACAGAGCAAC TGGAAAGCTC GACTACTCGG AAGAAAGTCT CTATGCAGTA
GAAGAAATGG CAGCGGAAGC AGCTCAATAC AAAGATCAAT TAGATCCAGC
CACTGTAGAC TCGCTTACTC AAGTTCTTGG AAGCTATATT CTTGAGGTTG
CACATAGAAA GCATGGCGGC TCTTACGTTT GGCTTGAATC TGAAAACTCA
CCTGCCTTGG TAGTTGGTGA ACCAGAGTAC AGGCTAGCAC TCTCAACCTT
CGCCAAGGTA CATGGCCGAC TTTCTGGCGA CGAAGCAGAT AATCTTATTT
TCTTCTATCA AGGCTTTTCT GAAAGGCTTA AATCACCATC TCCCGGCATG
AGCGCACTCT ACAAATGAAA CCCGAGTTTG GGGGCCCAAC AATGCGCTCA
ACTGCCGCTC ACTTCGTTCG CTGGACGTCC AAAAGCTACG CTTTTGGCCG
CCCGTTAGCT TAATCGTTAT GCACAATAAA ACATGAAGAC AGCACTCATA
TTTGTAGCTC TAATCTTTCT CTCTGGATGT GACAACTATC AGTCATGCCC
TATAACTGGA AAATGGAAAT CCAACGAAAA GCTAACTTTA GAAAGCATGA
ATGAAACCGG CAGGATAACG GCAAAGCAAA GAGAGATTTT TGAGAACGGC
TTCTTTGGAA AACTAGAATT AGACATAAAT TGCAGTAGCT TCACAACAAT
ACTTGACGGC GTTACCGAAA CCTTTAATTA CGAGATAGTT CGCCAAACAA
AAGATTCCGT CACCGTTAGC TATTACAGCA AAGCGCTGCA AAAACAAGTT
GAGGTCACAT CTATTATCAA CGGAAATTGT TACTCGACAC CTATAGAGCA
GTTAAATTTC AATGAGTATT TCTGCAGAGT CGAGTAGCGC ATAACAATTG
ATTCAAGTCG TTCGCTTCGC TCACTGCGGG ACCGGCTAAA GCCGGCCCCT
TAACCAAACG TTAGGCAAAG GCTCAATGGA TCCATATTC CATAACATCC
ATAGAAACGA CAAAGAGATT GAGGGCGCTC ATCAACAATG CTCGAGCACA
ATCAATCACT TCATTGAGAT GGTCAAAAAA GGGGGCGAGC CCACCTATAT
GGCAAAGCTA CGTTTTCTTG ACCCTGACAA GTCTGAAAAA GAAGGTAAGA
ATCATATTTT TTATTTGTGG TTATCTGAAG TGCTGTACCA CCCTGCAACA
AATTTACTTT CTGGGGTATT TTTTGAAATC CCTGAAGGCT TTGAAAAGTG
GCACCAAATA GGCCAGCGCC TAGGCTTTGA TCCAGAAGAT GTCTTTGATT
GGATGGTAAT CGACAAAGGT CATGCTAAGG GTGCATACAC ACTAAAGGTA
TCGCGAGAGC GCTTAACCAC CGAGCAAGAA AGAAAGATT TTGACCGCTA
TATTGGTGTG GCGTCATATG AGTAGCCTAA AATTAAGCGC TCACGCCTCA
GCCTAACTAC TGGTTCAAGT CACTCGCTTC GCTCGTTCGG GACCGCGTTC
CGCGGCCCCT TAACCAAACG TTAGGCGCAA GGGCAATATT GGTCTTCAGC
ACCGAGTCAG GAAACACAAT CACCGAATCA GCGCGGTGTT CCTGAATCGA
ATGGTCGCTG ACAGTTGAGG CCGTTATTTG TGGCCAGCAA AGGAGTTGCT
TTCAGAGAAT GTGCACGTCA CAAATAACTT CCGGGGCCAA AACCGAAACG
CCGTGCGCTC CGCCGGTTAA GCTCGGCGCT GGCATCATTT TCGGCGCTCG
GCTGGGCAAT CTAACAATTG GCTCAAGTCG TTCGCTTCGC TCACTCGGGA
CCGGCGAAGC CGGCCCCTTA GCCAAACGTT ATGCGAGCCA CCATGAATAG
CGAAGAATTA TACAAAAAGG CTATGGAGTT AGAGTCCAAA TGCGAGCATA
AAGCGGCAAT TTCAACTTAC AAAGAAATTG TTAAGAAATC TAACGATCCT
CGACACTTCA TCGCATTCGG AGTTTGCCTC CAAAAATGTG GTCACTGGAA
GCAATCCATC GAGGTATTAG AATCAGGAAT TGCACTGAAG CCTCACTATT
GCGAGGGTGA TGCTCGTCTA TTTTTAGCAA AAGCACTTTT TAAATCAGGC
AAAAAAGGCC TTGCGATAAA GCAATGGCAA CATGTATCAA AAATGCAACC
TGAGTACCCA AGTTATGAGT CTGTGCAAAA TGAAGCCAAG AAAATGCTTG
```

```
CACAAAACGC ATAACAATTG GCTCAAGCCG CTCGCTCCGC TCACTCGGAC
GTCCGTAAGC TACGCTTCCG GCCGCCCCTT AGCCAAACGT TAGGCACCAC
ATGCCCTCCA TCAAGTCAGC AAGCCAATAC CAGCGCGCAA CATCGCTCAT
CTTCTTAGTG TCAGGCGCTG CTTGGCTATT CATCGTGCAG TCTTCGTTGC
TGCCATTGAC GGATGTCGCC CGCCAAGAAA TGGTTTGCCT TAATATCGTT
CTTGGTATTG CCTGTTTTGT TATAGGTAGT GCGGCAAAGC GTCAGCGAGA
ATTTCGCTGC CCTGACTGTG GGAACGAAGT AGATCAGAGC TTACCTACAG
AGGGTGATGG CGCCCCACTC CTAAGGCTGT GCAAGCACTG CGATATTCTA
TGGAATGTTG GCAAGACCCC AGACAGTTAA AGTTACCGCC TAACAAATGG
TTCAAGTCGT TCGCTTCGCT CACTCGGGAC CGGCTAAAGC CGGCCCCTTA
ACCAAACGTT AGGCAACAGG GGGTGACATG ACGCAATGTC CAAGGTGCCA
GCGCAATCTC GCAGCTGACG AGTTCTATGC TGGCTCTAGC AAAATGTGCA
AGGGTTGCAT GACTTGGCAA AACCTAAGCT ACAACGCGAA TAAGGAAGGT
CATGCCAACA CCTTCACCAA AGCGACATTT TTGGCGTGGT ACGGCTTATC
AGCACAGCGG CATTGTGGGT ATTGCGGTAT ATCGGAGGCA GGTTTTACAT
CCTTGCACAG GACTAATCCA CGCGGCTACC ACATACAGTG TTTGGGTGTT
GATCGCTCAG ATTCGTTCGA AGGCTATTCA CCTCAAAACG CTCGGCTCGC
CTGTTTTATA TGCAACAGGA TAAAATCAAA CATCTTCAGC GCCAGTGAGA
TGGACGTTCT AGGTGAGGCC ATTTCAAAAG CGTGGCATGG TCGAGGAATT
GCCTAACTAT TCAGTCAAGC GGACGCAAAC CCCGCTGCGC GGTCTTTGCG
CCGCTTATCT CAAGCGTTAG ATGAATAAAA GCCTCCACAC ATAGCCAGCT
TTACGGGAAC GAAGTTGATG CGACGCCTCT TTACTGCCTT AGTATTAATT
TTGATGATTT CTGGGTGCTC CTCCACATCA AAAACTGAAA GCCATAAACA
GCCGCCGAAC AATTCAAGCG ACACGACCGC CATACTGAAA TATATATTTA
CTGTTACCCA TGGCATAGAA GAGGGCTTTG CCAAGTCACT TGAGCAAGGA
AGTTACACAC CAGATGAATA TATGCTCATG CAGCAGGCTT TTAGCAATCT
TGATCTAAAC AGATTAACCA CTCTTTTATC ACCCACTTTA GATAAAAGCA
TGAACACCGC AGACGTAAAA CATTTCATGA TTTTTATAAA ATCTACTGCA
GGTAAGAACT TGCTAACAGC AGGAGAGTCT AGCACTTCTT TTTCCGGCAC
AATGGATAGA GTTCGATCAC TACCCTCCGA ACAGCAGTCG AAAATAAACG
AGTTCTTCCA TGCCAGCTAC ACAAAAACA CTTTAACAGC CATGGGGGCT
CCAGAGGCGG TACGCATTGT TTATGCATTT GGAGTGGAAT CCATGTGCAA
TTACGCTATG CGCAATAATT TCGAACTGTA TATTTCTATT ATAGAAAAAG
GCAAATGCCA ATAAACCACA AAAAACAAAG GCCGGATCGA TCCATGTGAA
CATCAGCGTT ACATCTAACA TGTGGTTCAA GCCGCTCGCT TCGCTCACTC
GGGACCGGCT AAAGCCGGCC CCTTAACCAA ACGTTAGAGG ATTACATGCC
ATCACTCCAA GAACTCCAAT CGCCGATTGA CTCAGCAATC GTGAACTCCA
TGATTGAGAG CACTCCCGAA ACATGAAGCC AGATTATTCT AACGTTGGTA
CGCGAATCTA ATTCCTCTTG TGTAGGTAAC TTTACACATG AGTTATCCAG
TCCTGAGGGG CATGCACCAG TTGGTCCGGC AGAGAGCTTA TTTGAAGACA
CTTACCAACT CGATGAGTTA TTCTACAGCC ATGGTGAGCG CTTATTCACC
AAGGCAATTT ATCGGGCTAA TGCGGTTGGG GACGGTTGGT CATATCACGC
TGAGTTTGAA TATGCGTAGC ATCCCTCTAA CAATTGGTTC AAGCCGCTCG
CTTCGCTCGC TCGGGATCGG CGAAGCCGGC ACCTTAACCA AACGTTAGAG
ATGGTCATGA ATAAACGCGC ACTAACCTTC GGGCTACTCA TAGCAATTCT
AGCTAGTATC ATTTCACAAG CGCTGCTTCA TGGCCAAAAG GTAATCGCTT
CCGACGTGGC TTTATATACT CCATATTTCC TATCTCCAAT ATTTAGCATG
CTGATACAAG CCACATCGAT GCTGGCATGG GCCATACCTG GACTTTATGT
AGGCTACTTA TGCAAGAACA AGCCAGCACA ACATGGAGCA AAATTGGGGG
CAGCATATGG AATACTTCTT GGATTAATTG TATTCGCAAT GCGAGCTTCG
ACCCAATTAA CGTAGATTCT AAGTTAATAA TCGCAACATC TGCTTTAACA
CAAAAAGCAA AATATTCAGT.GCACTTTGCG CTAGTTGCTC CTGCCGGCTA
TCTTCTTGCA AAGCATCGTG CAAATCTCTA ACAATTGGTT CAGATCGTTC
GCTTCGCTCA CTGCGGGACC GGCTGAAGCC GGCCCCTTAA CCAAACGTTA
GGCAACTGAA TGATCACCTG CATTCCGGCA CGTGAATTCC TGCGTAAAGT
ATGCGGCCTG TACGAAGCCT CAACTAATGT AGTTAAGTTG CGTGTATGGG
```

```
CTTGTGGATA TGGCATCGCA ATGGATCTAA CTGTCAAAGG TAAATCTGTC
CTTTGTGCGG TTGCGGGAGT ACTCCGCCAA GAGGTCGAAT GCTTTGCTCA
AATTGGGCTT CCGAACGTAA TTCAGTTAGT AGGCGACAAG GCGTCAGAGA
ATCAACTAAA GCTCATAGGC ATGGAACCAC CAATCGAACT TCATATCTCT
CGCGAACAGA GCAGGCTCCA AGTTGTAATC TTGTACGAGG GTCAGGTAAA
GGCTACATAT GTGCTGTCAG CCGCCTAACT ACTGGTTCAA GTCGTTCGCT
TCGCTCACTC GGGACCGGCT AAAGCCGGCC CCTTAACCAA ACGTTAGGCT
TTCAATGAAA ACAGTTCCAG TGAAAATATC AGAAGTCGAA CTAATAGAGA
GTTTTGGGAA ATTCCTGATC AATCAAGACT TAATCGACTA TGAAAATTCC
CACTTCAGTG GCGACGACAA CCATAATGCA GATGTAGCCT TATCTTTAAA
GCCAGGGAAA TGGCCAGGCA TTCAAGTCGA TAAACTACAC ATAGAAGTAA
AGTCACACCA CTCAGAAGAC TCTCAAAACA CCATCAACAA AATATTCGGC
CAATTACTAA AAGAAACCGG AAAGCGAAGC CTCGATAAAG AGAAAGAGTG
CTTAGCTATA TTGTTCCCTT ACGAGCGCGG CGCATGGCCA GGTCGAAACA
ACAAAACAGT AACAAGAATT GAAGGTGAAG CTTATTACCG GAGGGGCTTT
TCGAGAATCG ACAAACAGAC GTTTGTTAAA TTTGGTGACT TGGTCGGTGC
CAAATACATC CTTTCCTTTT CTACAGCATC AAACACATTG AACGTATTTG
AATGGAAAAA TTTCTTAGAT GAGGAATTCA GCCCGATGAT CAGCCTAACA
AATGGTTCAA GCCGTTCGCT TCGCTCACTC GGGACCGGCT AAAGCCGGCC
CCTTAACCAA ACGTTAGACG CACCGGAAAT TTTGCATGGG AAACCAGAAA
ATGGATTTGC AGATAAACGA TACAAAGGTT GAGTGGGTTT CTCCAATACT
GAAGCAATGG ATCAGCATCA ACAAAGAATA CGTCAAGCAA TATGATTTCA
AAGACTGCCT GCACTGGTAT AACGAAAGGG CAAATATAAG TGCTTTTGCT
GGTGCCGTTT GGAAGTCTGG AGGTTTTGCG CTGGAAGAAT ATTCAACTAA
AAAAGGCACC GAAGAAAACA GAGCCAATGG TCGTGTCGAC CTATATTTCT
CCAATGACAA CGAGCAAGCC ATTGTTGAAG CAAAAAATGGA ATGGCTCTAC
TTCGGAAAGC GCACAAGACT AGATTTCAAA GAAAAAATAG ATCGTGTAGT
TGAAAAAGCA AAGAATGACA TAATTAACAG CCTGCATGCC AACCCCTACG
ATCTAGGGCT TGGGCTTTCC TTTATTTGCA CATACTGGAA AAAGGGTTAT
GACGCATCCG CCGACATGCA AGCCCTTAGA GCGCTTATGC AAAATTATAA
CTGCGCATTT TATGCAATTT TTGAAAACAG CCCCGACAAC GAAATTGTTA
GCTCAAAGGG CAATATCTGC AACGCTGTGA TTTTAGTTGG GACGGCGCAC
AGCTGAATCG TGTGTGTGCG TCTAACAATG CGCTCAAAGC GCTCACTTCG
TTCGCTGGGA TCGGCTAAAG CCGGCCCCTT AGCTTAATCG TTAGCACTAG
GACTTCCGAC CATCATGAGT GATAGAGACG AATTTTCTGC CCCAACAAAA
AGAGCGCTAG CCGAAAGGAG TGGCTTTAGG TGTTCTTATC TTGGTTGCTC
TAATGCAACC ATAGGGCCTA GTGAAGAATC AGAAACAGCC GTAGCAAGAA
CGGGGGTGGC GTGTCATATA ACTGCCGCAG CGCCCGGCGG AAAAAGGTAT
GACCCAACAT TAAGCCCTAC GGAACGAAGC TCAATCTCGA ATGGTATATG
GATGTGCCAA ACGCATTCAG TTGAAATAGA TAGAGATGAG GCCCGATACA
CATCGACCTT ATTAAATCAC TGGAAAAATA TATCCGAGAG CCGAGCAGAT
TATGCAAAAA ATCATGGCTG GGATATTTTT GACAAATACC CCTTCCTTCA
TATTGACTCG CTAGCCAACA TAGACCTGGC TCTTACCAAA AGCCCTTCCT
CAAATAGCCT TATCGGGAAT GCCATTACAG ACAGCTGCCT CCCTCAACTA
TGGGGTAAAG AGCAATCTGT AATCATCAGA GACCTAATAA TAGAACTTTA
TCGAAATGCC TTCGATCACG GCGAGGCTAG CTCATTCGAA ATATCCATAT
CGGAGCAAAA ACTAGAAATA GTTTACGATG GCAAAAAATT TGACATCTTC
CAACTTCTTG ACCACCAGAA TGCAAACGGT GGCGCCGATA CCTTGCAAGA
AATTGTAGAA AAATATGGCA GTAACTTTGT AGTCAACTAT AGCCACGAAG
GCAACAATAA AATAATAATT CACAGGCTCT CTGACTTTTA CGCGCTTGCA
CCATCCCTCC CGTGCGTAAT ATCACTGAGT GAATACGATG ACAAGGCCCT
AGAGTTAGAC CTGGCTATTT ATGAGCGCTG CGGTGCACTG TACATAATTC
TACCGTTGCA TTTTTGTAGA TCAGATGTCA GGGGGCTAGA GTCGCAGCTA
GCCGCCTTTG AACCTAATGG AAAGCCAGTT TACATTGTAG GCTCAGATGT
GGCAGAGCCT ACAAGAAAAG CAATTATAGA CAGGCTTCCC AACTTCACGT
TCGTCCAAAA GCAATGCTAA CAATGCGCTT AACTGTCGCT CACTTCGTTC
```

```
GCTGGATAGT CAAAAGCTGC GCTTTTGTCT GTCCGTTAGC TTAATCGTTA
GGCGCAAGGA GGGACCGTGA CTGAAACTGA GAAAATGGTG GGTAAGTTCG
TCAGCGGTTT TGGCGGGCAG AGATACCGAG AAATTTTTGA AGTCCTCGAA
TCCAGTAACC TTCGCCCACT GGGCAAGTCA AATACTGAAA CATTGCTATT
TCAGCTTCGA GGGGCTGATA GTGAAATGCT AGATATTTTT GCCTTTCGCT
TGGGGCCGCC GCCAGTAATT TCGTTTCCCA AATCATATTG GCTAGGTCGC
CCCAGTGAAT TAAGCGCTCA TCTATCCAAT TTTTCATTCT CGGAAAAGCC
AGCCATAACA GGCCCGGTTT CTGACTCACA GTATTCGGCA GGCCAGGTGG
AAATCACCCG CTCTACTCAT GAGAGGATTA TTGAGGTTTG CAACCGTGTC
TGTGCTTCCC TGCAATAAGC GCCTAACAAC TGGTTCAAAT CGCTCGCTCC
GCTCGCTGGG ACCGGCATAG CCGGCCCCTT AACCAAGCGT TAGATGCAAA
TAACTTGAGG GGCACATGCA AGACTTTGGG TCCAGACGAA ATGCATCATT
AGAGGACAGG GCTGCGGCTG AGTCTGTTAT TGAACGTGTT TATCTTGCGA
TACAGCAGCT TTGCACAGAG ACTGGTGACG TAAGAAATCG GCTTCAAATA
GCCGTTATGA CTCTATTGCC CCTTCAGGCG CGTAACTTCC CCATTGCGTT
GCAGCAAGAC TTCGATTGGA TTGTCAGAGA ATCAACCAAA TACAAATCAC
CATATCCGCA GTTTCGGGGC GACCTTGAAG CAACGATGAT GCGAATAAGG
AACTCAACTG GGCAAAAAAT CGCGCAAAGA ATTTTCAATA TTTACTCGTC
GCTACAAGAC ATTCGAGGTT TTCCCCTGCT TGAATACAGG GCAATAGATG
AGTAAGCATC TAACAATTGG TTAAAACCGT TCGCTTCGCT CACTGGGACC
GGCTAAAGCC GGCCCCTTAA CCAAACGTTA GGTAACCAAG GGAAATTCAC
TTGAGTTGTT ATGTATTGGG CACAAACAAC CGCCATTAAA GGACGGTTTT
ATAGTAAATT TCATCGGACT GTTGAACTAA AATGCTTATA CGCTTTGCTC
TACTACTTGC TGTTATGCTC CTCGCTGCAT GCTCGTCAAA GCAAAATCCA
ACGCCGAAGT GTACTGCCAG CGTCCCCCCG CCCTCTTTAC CCGAAACATC
CACAGTATGC CTAGGGGAAA GATGTAATTG GGAGGTGCTA TTTCCGTCAG
GAAAATACCC TGCATCCACA GAAGGCTGCA GGGCGCCTGT GGTGCAGAAC
CAGCCTTCTT CCTACCCGCG AGAAGCACTT GATCAGTGTA TTGAGGGGTA
CGCTTGGGTA GCGGTTTTTC TGAATGCCGA CGGGGTCCAA ACATCAGCAA
AGGTACTTCA ATCATCGAAT AAAATTTTCG ACAGAAATGC CTTGCTACAG
GCCAGTAATA TATTTTTTGA GCCTATGAAA TGTCAATCCG AGCGTTATGA
TTCCGTTGTT CTGATGCCAT TAAACTACCG CATACTCCCC TAGTAGCGGG
ATTGATCCTT ACAAAATTCA CTACTTACGT CCAAGTTGAA GTAGGCAGTT
TAACAACTGG TTCAAGCCGC TCGCTTCGCT CACTCGGGAC CGGCTAAATT
CGGCCCCTTA GGCAAACGTT AACTATCAGA AGGGCGGTTG ATGTCAAGAT
TTGCGCTCGC GTTGATTCAC GGAGTACCAA CGGGTTTTCT TGTCATTTGT
ACTTTGTTTG TCTGTTTCAT CTACCTCAAC CGATTCGAGA AAGTTGGAGG
ATACTCAGAC GGGTGGGGTT TTGTTGGAAG AGTTGTCTGC GCATCTATAG
CTATGGTTTT CGTGTCCGCA GTTGGCCATC TTCTTATTGA AGCGGCAGTC
AACTGGGGGC TGCAGCAGCT TGGTTATGAG CTGCCAAACT ATGAAAAAAG
AAGGACTTGT AGTAGCTGCA AGCCGAGCAC TCCAGGTGAC TACATGTTCG
GCTTGCTCCT CGGGGGTGTG CTTGGCGCCG GCTCGGCAAT TTGGCTCTGG
ACGCGCCTGG CGCTCCGATA TGCGCTGTTT CGCGGCGAAA ACTGATAGCT
GAACCTTCCA TCGAGGAGAT GCAAAAGCGC TGCTGCGCGC CATCTACAAA
GACCCGAAGC ACCTCATCCA GGCGCTCTCA GCCCGAGCCT GACTGGCTGT
GGCTATCAAC ACCTCTTCGA TACCACTACC CGCCAGAAAC GACAAAGCCC
TGCAAAAAGC AGGGCTTTGT CTTTGGGGAT CTGGAGCGGG CGAAGGGAAT
CGAACCCTCG TCATGAGCTT GGGAAGCTCA GGTAATGCCA TTATACGACG
CCCGCTCGGG CGGCTGACTT TTTACCAGAA TCGCCCGGGA AGGTGAAGCC
GGGCGCGCGT CTTGCGCCCG TTTTATTGCC GGGCGCTTCA TAGCGCCACG
GCCCGTGGCT CTCGTTCCAC GCTGCGTGCG TGGCCCTGCG TGGGTGCCAG
CAGGAAGGCC AGCAGGGCAT CGCGGGTCTG CATCCAGGCG GCCTTGTGTT
CCATGTCGAG GAAGTGGCCG GCCTGGGCGA TGGTGCGGAA CTCGCAGTGG
CGCACGTACT GGGTGAACAG GCGCGCGTCA GCCGGGGTGG TGTACTCGTC
CCACTCGCCG TTGACGAACA GCAGCGGTAT CTCGATCTGC CCGGCGAAGC
TGACGCAGGA GCGCCCGCCG TTGTTCAGCA CGGTTTCCAC GTGGTGACTC
```

```
ATTTGCTCAT ATTCATAGCG CTCCAGGCCG GTGACGTGTC GATGGTTGTA
GCGCTTGAAC AGCGAGGGCA GGTGCTTGCC GATGGTGCCG TTGAGCACCA
TGCCGATGCT CTCGCGGTCG CACTCGCGCA TCACCACCAG GCCGGCGCGC
AGGTAGCCGA GCATGGCGCT GTTGACGATC GGCGAGAAGG AGTTGATCAC
CGCACGCTCG ATCCGCGATG GACGCCGGGC CAGCGCCTGG AGGGTGGCGA
TGCCGCCCCA GGAGAAAGGA CAGCACGCTG TTCGCAGGCG AAAATGTTCG
ACCAGCTCCA GGAAGATGTC GGCTTCTTTC CTCGCSGCTG AAG
```

Figure 3B-1
SEQ ID NO:2

```
AAGCTTCTGG TACGAACCTG GGGGCGCTCC GGCACGCACA AGGGCATCGA
CATCTTCGCC CGCCAGGGCA CCCCGGTGCT CGCCCCCAGC TACGGCATCG
TGGTGTTTCG CGACGAGCTC GACATGGGCG GCAAGGTACT GCTGATGCTC
GGCCCCAAAT GGCGCCTGCA CTACTTCGCC CACCTCGACA GCTACAGCGC
CCTGCCCGGC CAACCCGTAC TTCCCGGCGC CCCACTCGGC ACGGTAGGCA
GCACCGGCAA CGCCCAGGGC AAGCCGCCCC ATCTGCACTA CTCGATCGTC
ACCCTGTTGC CCTATCCCTG GCGCTGGGAC AACAGCACTC AGGGCTGGAA
GAAAATGTTC TACCTCGACC CCACGCCAAT GCTGAACGAA GCGGCAGTAG
ACAGCCGAAA AACCAGCCAG TAGCGTCGCA GGGGAATGCA CCACCGGTCT
TGCCCGATCC GCCTGTCCTT TTACCAATCG CAGAAGAGTC GCTTTTGTCG
AATCGCCTGT GAGGAAAAAC AAGGACTTGC TGGACGACAA GGAACGTTAT
GCGACACAAG TGGCGGAATA AATTACGCCA TTTGTGTCGT CTACTTATAG
TTATATGCTG ATCTAGATAT GAAGTACAAA AACATAAAAT CAGCAATCCA
CAATTTCGGG CACAGCTTTG TAAGCTCAGT GAACTATGTT GACCATGATT
TCGTTGCCGA CGAAATTGGG AAGATTCACA AGAAAGGCTA TGATATTGAA
ATAAACTGGC TTACAAGGGA GTTCAAGCCC GCTCAGCTTG AGTCAGAGAG
AATAAAAAAA TCAATTGGTT ATTGGGGTGA CAACCTAAAG AAACATTGTG
CATCCCATAG CGTAAATCTG GAAAATCTAT GTTCTTTATC GTTTATCTGG
CCGACAGGTC AAAGTAAATA CATGCATGCC ATTGACGACA AAGGCACAGA
ACACAAAATT TACATCAATG AAGCGCAGTG ATACGCATAT AACAATTGGT
TCAAACCGTT CGCTGCGCTC ACTGGGACGG GCTAAAGCCC GCCCCTTAAC
CAAACGTTAT GCGAGCCACC ATGAATAGCG AAGAATTATA CAAAAAGGCT
ATGGAGTTAG AGTCCAAATG CGAGCATAAA GCGGCAATTT CAACTTACAA
AGAAATTGTT AAGAAATCTA ACGATCCTCG ACACTTCATC GCATTCGGAG
TTTGCCTCCA AAAATGTGGT CACTGGAAGC AATCCATCGA GGTATTAGAA
TCAGGAATTG CACTGAAGCC TCACTATTGC GAGGGTGATG CTCGTCTATT
TTTAGCAAAA GCACTTTTTA AATCAGGCAA AAAAGGCCTT GCGATAAAGC
AATGGCAACA TGTATCAAAA ATGCAACCTG AGTACCCAAG TTATGAGTCT
GTGCAAAATG AAGCCAAGAA AATGCTTGCA CAAAACGCAT AACAATTGGC
TCAAGCCGCT CGCTCCGCTC ACTCGGACGT CCGTAAGCTA CGCTTCCGGC
CGCCCCTTAG CCAAACGTTA GGGGCCAAGA TGGATCTTCG CCAGACAAAG
CCAATACTAG TTACAGTCTT AGCCACTGCC TTGGTGCCAT GGTTTTTGG
CTGGTATGCG TATTGGGAAA ATCCTCAAGG CATACTTTTG TACACTCCGG
TGGCCGGCCA TCCCCATCCT CAGGGCTCTC CAGCATTTCC TATTGGAGTA
ATGGTTGGGC TGGCCGCTTC ATTTCTGCTC TCTTTGCTTT TTGTAGGCCT
AGGGGGAATC GCTGCATACA TAGCAAGTTC AGTGAGCTCA AAGGCTAGGG
CTAAGCTGTT TTGCAAAATC GCAGTCACAT CCCTGGCTAC TTCAACTATA
GGAGCTGCAG TCTATGCAAT GCTCCCCTAA CAAATGGTTC AAAGCCGTTC
GCTTCGCTCA CTCGGGACCG GCTAAAGCCG GCCCCTTAAC CAAACGTTAG
GCAGCACATA TGACTCGTTC GTGCCTATAC ATGTTTATCG CCTCAGCCTT
GATAGCGTGC GGCGATCCAC CTCTATTGGT TACGCCACTG CCAAATGGCT
ACAATTTCCA TTCCAACGGC GGGGAGTTTG GCTACATCAA GAATCCAGAT
GGATTAAGGC TCGCCGAGTA CTTTGGTATT CGTAATGATG GTCGCGAAAC
CTGGTGCACT GACTTTTCAT GGGAAAGCGA TATCGTCATT TGTAAGCTTA
TTGAATATAG CCAGCATGGA TTTGACGCAT CGCATACAGA GTTTTCTGTA
CTTGACACAA AAACTAGCGA GGTTAGGGTA TTTCCCGATC AAGCGTCTGC
TCAAAATTTC TGGGCCGCAC GCTTTAATTC AGGACTACCT CAGCTTCACC
GGCACTACCC TTCAACCTCA GAGAAGTAAT ATTTTGTGTG TCAGTGCAGC
CTAACAATGC GCTCAACTGC CGCTCACTTC GTTCGCTGGA CAGTCAAAAG
CTGCGCTTTT GCCTGCCCGT TAGCTTAATC GTTAGAGGCT TATTTAGCTC
ATGCGCATAG ACATAGACTT TTCAATATTC ACGCTCGCAC CGTCGACCGA
AGGCGTAATA TCAGGAAAAA TCGAGGTCAG TGAACTACCT AGAACTGGCG
```

```
AGATAATTTC ATTCTCCTTT GCGCCAAACA AGTCTAAATT CCCGGCAGAG
CCAAGATTCA ACCCGTTGCT TAAAGTTGAG AGAGTGATTC ATAGCGTAAA
TGGTCAGAGT CCAGCTCTTC AGTTAGAGAA TCTGATGCTA CCAAACAGAG
AAAGTGTCGC TGAAGTCACT GCTTTCCTAG AGCAAGGCTT TGGCCTATTT
TTCAGCCCAA CCGGTGAGTA ATCCTCTAAC AATGCGCTCA AATCGCTCAC
TACGTTCGCT GGGACCGGCT AAAGCCGGCC CCTTAGCTTA ATCGTTAGAG
GTCAGCACAT GGCAGTGCAG CAACTCGGGC CAACCACAGT ATCCGTAACC
GAATTTGCAT GGGACGGAAG CGATCTTGGA AATACTGAGG CCAATGAATT
CTGGTCACAG CTCTCTGCTC AGCTTCAAAA AATAGCTATC TCTGAGTTTT
TAGCTGGCAA TCGCCCCAGC AGCATTCTTC GCAACGACCC ACGAAACATT
ATTGTTCTCT CATTTTCGGC GCCGCCAAAG TTCATTAAAA TCAACCACTG
GCTCTCTGCG TGTACACACA GAATTTCAAC ACGGAAATTA CTGCTACGAC
GGAAACGGCC TGTACTTACG AAAATTTAGA GTCTGGCGAC TTTCTTGCAT
TCGACACAGC GGCGTTGGTG CATGCCCTCT AACAATTGGT TCAAGTCGTT
CGCTTCGCTC ACTGCGGGAC CGGCTAAAGC CGGCCCCTTA ACCAAACGTT
AGGGCACCGC GCATGAGAAA TGAAGACGGA ACCTTTTGCA AAGACTGCCA
CCATCAACTT GATGAAACAC TAGCATCTAG CGCAAATTAC TCATGCCCCA
ACTGCGGCTC CACAAAAAAA TACATGAACA TGTCCATCAC TGATGGAATT
GGCCTATACG ACTCTTTGGG TGCCCAAGCT AAAGATCCAA GTTACCCGGC
AAAAGAAAAA TCAGATGGGA AACATTTGTT GGCTGGGAAC GCAGTCATAA
ACTGCAAAAA ATGGTTTACA AGACAAGAAC TATCGATCGA ACCAATGACG
CATACCAAGA AATAGTAGTC GACCTTAAAA CAGGGGGAAT AATTCATCAC
TGTGAAGAGC CACTTTCAGA GCAYTTKGGC CATGGCACCG CAAAACCAAA
GCCCTAACAA ATGGTTCAAG TCACTCGCTT CGCTCGTTCG GGACCGGCTA
AAGCCGGCCC CTTAACCAAA CGTTAGAGGT TACCTGTGAC AGATTCGCGC
CCGTTACTGA TCCCTGCCTC GCAATATGAT ACGAGCGTTC TTCTCGCCGA
ATGGCAATGG CTCACCCCCA AAACGGATAC GCCACTTTTT ATTTCCATAT
TCGGAGACTG GGTATTTGGC AACCCCAATG GAAGTTTGTG GGTTCTTTCA
CTCCTAAAAG GCACTTACGA GCAAGTAGCC GCAAACTCTA ACGAGTACAA
CACCCTCAAC AAATCGGCGG AGTGGATTGA TCAAACATTC ATCGCCAGTT
GGCAGTCTAT TGCCGCAGGC CATGGGTTAA TCCCAGAACC AAACCAATGC
CTCGGCTGGA AGGTTCACCC ATTATTAGGT GGAAGTTTTG AGCCAGCCAA
TCTCCAACTC TTCAACATGT CGGTGTATCA ATCGCTTATG GGTCAACTTC
ATCGACAGCT TAGCCAAAAA CAAACCCCGG CAAGTAAAAA ACCATGGTTC
CAGTTCTGGT AACCTCTAAC AACTGGTTCA AGTCGTTCGC TTCGCTCACT
GCGGGACCGG CTAAAGCCGG CCCCTTAACC AAACGTTAGG CGCAAGGGCA
ATATTGGTTA TTCAGCACCG AGCCAGGGAA CACAATCACC GCATCAGCGC
AGTGTTCCTG AATCGAATGG TCGCCTGACA GTAGAGGCCG TTATTTGTGG
CCAGCAAAGG AGTTGCTTTC AAAGAATGTA CACGTCACAA ATAACTTCCG
GGGCCAAAAC CGACACGCCG TGCGCACCGT CGGTCAAGCG CAGCGCTGGC
CTCACTTGCA GCGTACGGCT GGGCAATCTA ACAATTGGCT CAAGTCGTTC
GCTTCGCTCA CTCGGGACGT CCAATAGCTG CGCTATTGGC CGCCCCTTAG
CCAAACGTTA GGCCAACATA CTCAACGCAT GAAAACAAAA TATCACATAA
ATATAATTAT ATTTCTCGAA ATCATAATTC CTTTAGCACC AATAATTTGG
GCAATTTTCA CTCAGTCAAG CCCCGGCTTT GGCCCAACCC TTATATCAAT
GCTCATCCTG CACATCGTCG GACGAATAAT TAGCCGAAGC ATCCCTGCCA
GCTGTGACTC ATGTGCTGAA AAAATAAAAC CCAAAGGAAC CTCCGCAATC
TACTACAACT GTCAAAAGTG TGGATTTAAA TACTCAAAAA CACTTAACAG
CAGCAAAAAC TTCCATAACC ACTAACCAGA AAATACTAA GGCGCCATCA
TGTTATAAGC GCCGTAAGCA CTAAAGACTT GTACAAGCCT AACAACTGGT
TCAAGTCGTT CGCTTCGCTC ACTGCGGGAC CGGCTAAAGC CGGCCCCTTA
ACCAAACGTT AGGGCACTCA ATGCATCGCT TCCTAGCCAC ATGCCTACTA
GCTACATCTA TTAGGCATA CGCAGAACCT GAAAATAATA TCGACTGCAG
CAACGCATTC TCAACGCCGG ACATTGAACA TTGCGCATCA ATCTCTCTTG
AGAAAACAGA GAAAGAGCTA AATTTAGCAT ATCAAAAATT AGTCAAAGAC
CTTTCTCAGC CAAACAATGA ATACGAAAAT TTCACCGAGT ACAGGAAAAA
```

```
ACTTTTAACG GCTCAAAGAG CATGGATCGC GTTCAGGGAA GCAAACTGTG
CCACTCAGTA CGAAATGCAC AGATCTGGCA CTATTCGCAA CAGCATCTAT
CTAGCCTGCA AAGAAAAGCG TGCCAAGCAG CGAATAAAGG GAGCTTCAAA
ATTATGCTCC GTACTAGCCC TAACAAATGG TTCAAGTCGT TCCGCTTCGC
TCACTGCGGG ACCGGCTAAT GCCGGCCCCT TAACCAAACG TTAGGCCGAC
AATCGCAATT CCTAGGACTG CACGTGAACT GGATCCGCAA AATGTTTCGG
CGCACAGCAC TAGCGCCGCC CCAACATCGC GAGGACGAAG CTGTCAGTAC
AAGCCAAGAA GGAACGCCTC CCTTTCGTCA TTTGACAGTT GAGAATTCAT
GGGGAAGTTG AGGGCGGAGC TATTCCTTCA GTCACCACCC GAGAACATCC
TCAGAAGATC TGTTTCTCGT TTGGCGTGCC TAAGTTCGGA TGGTCAACGT
TCGAGATCCA TTTCGTCGGA AATGGCCACT TCATCTGCGG CATCTCTGAC
ACTCCAAATG ACTTCTACGG TGACTTGGCT ATCGCCCTGG CTGAGCAGAA
AAGTTCTTTT TCGGTAGCGG CGCACCTTGA GCCTGAGACC TTTGCCTTCT
ACATCGTTGA TTCGACAATG TACTTGTGCA AGTTCGATGA ATTCGACGAT
TATGAGTCCG CCGCCGAAAG CCACGAACAG TTGGTCTCCC ACAGCTTTAT
GTCCATTGAA GTATCTAGGG AGTACTTTCA GAAGTCTCTC AGGACCTTGG
CCGTCCAATG GCCGGATACG CCTTCAAGAG ACTGGGCGCA CCCATTTCCA
CGTGCGCAGA TTGAAGGCTG ACTGCCTAAC TATTCGCTCA AGCGGGCAGC
GTTAGGCGCC CTCATTCGGA GTCACGCTAT GGCAACCCGA GAAGAAACAG
AAGTAGCCAT TGCTGCTCTT CGCAGCGAAC TCAATGGCAA CGAATCGGAA
TACAGCTTTC ACATTCCCGG TTGGGCGCCA GAAACATCAG TCATGGGATT
TCGCTGGATG CAAAGCCAAC TGTGGGAAGG CTTCTACGTA AGCTATCGCG
TAGAGCACTC GGCCAAGCGC GTCGAATTCA AGTGCTGGGA GTACGGCGAG
CCCGAGCCGT CTTGGCTGCA AGTTGGCTAG GGGGCCGGCA AGATGCAATC
GCGGCGAGCG CCTAACACTG CAGTCAACCG GACACCAAAC TGTACGCAGT
TTGGTTCCCT CCGCTGCGCT CCGGTGCCGG TTACTTTCAA CGTTAGGCAA
CTCAGATGAG TGCTCCAGAC GCAGAACTTC TCGCATTGTT AGCCTACCGA
ATGGAAGCTA TTTCCATTGG GCATTTGGCA TTACGCCATC ACATGACGTG
GGACGAAACA CCTTCAATGG AGGTGTACTT CAATGGCATA CAAGTACTCG
AGGGAAAGGC CACGGGTTTC ACTAATGCAG CCATTGAGTC CGCAATTATT
CATTGCAGGG CAATCCTTGG AGTTTGTTGG GCTGCAGTCC TCCAGACACT
CTTCCACAGA AATTGCAGAG CGCACTCGAC GCAACAATCC CGATGACTAT
GGCATTGAAA GCTTCAATGG CTTATCAATG CTAACCAAGG AAAAAGCACT
AGCCTACTAC TCTGGCGAGC TGCCAGAAGC GGAAGTTGCT CTAGCGCTCA
TATTCCACTC AGCGAACAAA GGGCTTGCAC ACACTACAGT GTCCTTTACG
CGTGACAGTG GCGACGCCCA CCTGATGGAA ATTGCATTTC GCATCGTACC
AATCCTGCTT GTAAATGGCT TCTACGCTCC ACTGGAAATC ACGCCACCAA
AATATGAACT GATTTCACGC CCAAGAGTCG CCATAACAAA TGGTTCAAGT
```

Figure 3C-1
SEQ ID NO:3

```
AAGCTTCTGG TACGAACCTG GGGGCGCTCC GGCACGCACA AGGGCATCGA
CATCTTCGCC CGCCAGGGCA CCCCGGTGCT CGCCCCCAGC TACGGCATCG
TGGTGTTTCG CGACGAGCTC GACATGGGCG GCAAGGTACT GCTGATGCTC
GGCCCCAAAT GGCGCCTGCA CTACTTCGCC CACCTCGACA GCTACAGCGC
CCTGCCCGGC CAACCCGTAC TTCCCGGCGC CCCACTCGGC ACGGTAGGCA
GCACCGGCAA CGCCCAGGGC AAGCCGCCCC ATCTGCACTA CTCGATCGTC
ACCCTGTTGC CCTATCCCTG GCGCTGGGAC AACAGCACTC AGGGCTGGAA
GAAAATGTTC TACCTCGACC CCACGCCAAT GCTGAACGAA GCGGCAGTAG
ACAGCCGAAA AACCAGCCAG TAGCGTCGCA GGGGAATGCA CCACCGGTCT
TGCCCGATCC GCCTGTCCTT TTACCAATCG CAGAAGAGTC GCTTTTGTCG
AATCGCCTGT GAGGAAAAAC AAGGACTTGC TGGACGACAA GGAACGTTAT
GCGACACAAG TGGCGGAATA AATTACGCCA TTTGTGTCGT CTACTTATAG
TTATATGCTG ATCTAGATAT GAAGTACAAA AACATAAAAT CAGCAATCCA
CAATTTCGGG CACAGCTTTG TAAGCTCAGT GAACTATGTT GACCATGATT
TCGTTGCCGA CGAAATTGGG AAGATTCACA AGAAAGGCTA TGATATTGAA
ATAAACTGGC TTACAAGGGA GTTCAAGCCC GCTCAGCTTG AGTCAGAGAG
AATAAAAAAA TCAATTGGTT ATTGGGGTGA CAACCTAAAG AAACATTGTG
CATCCCATAG CGTAAATCTG GAAAATCTAT GTTCTTTATC GTTTATCTGG
CCGACAGGTC AAAGTAAATA CATGCATGCC ATTGACGACA AAGGCACAGA
ACACAAAATT TACATCAATG AAGCGCAGTG ATACGCATAT AACAATTGGT
TCAAACCGTT CGCTGCGCTC ACTGGGACGG GCTAAAGCCC GCCCCTTAAC
CAAACGTTAT GCGAGCCACC ATGAATAGCG AAGAATTATA CAAAAAGGCT
ATGGAGTTAG AGTCCAAATG CGAGCATAAA GCGGCAATTT CAACTTACAA
AGAAATTGTT AAGAAATCTA ACGATCCTCG ACACTTCATC GCATTCGGAG
TTTGCCTCCA AAAATGTGGT CACTGGAAGC AATCCATCGA GGTATTAGAA
TCAGGAATTG CACTGAAGCC TCACTATTGC GAGGGTGATG CTCGTCTATT
TTTAGCAAAA GCACTTTTTA AATCAGGCAA AAAAGGCCTT GCGATAAAGC
AATGGCAACA TGTATCAAAA ATGCAACCTG AGTACCCAAG TTATGAGTCT
GTGCAAAATG AAGCCAAGAA AATGCTTGCA CAAAACGCAT AACAATTGGC
TCAAGCCGCT CGCTCCGCTC ACTCGGACGT CCGTAAGCTA CGCTTCCGGC
CGCCCCTTAG CCAAACGTTA GGGGCCAAGA TGGATCTTCG CCAGACAAAG
CCAATACTAG TTACAGTCTT AGCCACTGCC TTGGTGCCAT TGGTTTTTGG
CTGGTATGCG TATTGGGAAA ATCCTCAAGG CATACTTTTG TACACTCCGG
TGGCCGGCCA TCCCCATCCT CAGGGCTCTC CAGCATTTCC TATTGGAGTA
ATGGTTGGGC TGGCCGCTTC ATTTCTGCTC TCTTTGCTTT TTGTAGGCCT
AGGGGGAATC GCTGCATACA TAGCAAGTTC AGTGAGCTCA AAGGCTAGGG
CTAAGCTGTT TTGCAAAATC GCAGTCACAT CCCTGGCTAC TTCAACTATA
GGAGCTGCAG TCTATGCAAT GCTCCCCTAA CAAATGGTTC AAAGCCGTTC
GCTTCGCTCA CTCGGGACCG GCTAAAGCCG GCCCCTTAAC CAAACGTTAG
GCAGCACATA TGACTCGTTC GTGCCTATAC ATGTTTATCG CCTCAGCCTT
GATAGCGTGC GGCGATCCAC CTCTATTGGT TACGCCACTG CCAAATGGCT
ACAATTTCCA TTCCAACGGC GGGGAGTTTG GCTACATCAA GAATCCAGAT
GGATTAAGGC TCGCCGAGTA CTTTGGTATT CGTAATGATG GTCGCGAAAC
CTGGTGCACT GACTTTTCAT GGGAAAGCGA TATCGTCATT TGTAAGCTTA
TTGAATATAG CCAGCATGGA TTTGACGCAT CGCATACAGA GTTTTCTGTA
CTTGACACAA AAACTAGCGA GGTTAGGGTA TTTCCCGATC AAGCGTCTGC
TCAAAATTTC TGGGCCGCAC GCTTTAATTC AGGACTACCT CAGCTTCACC
GGCACTACCC TTCAACCTCA GAGAAGTAAT ATTTTGTGTG TCAGTGCAGC
CTAACAATGC GCTCAACTGC CGCTCACTTC GTTCGCTGGA CAGTCAAAAG
CTGCGCTTTT GCCTGCCCGT TAGCTTAATC GTTAGAGGCT TATTTAGCTC
ATGCGCATAG ACATAGACTT TTCAATATTC ACGCTCGCAC CGTCGACCGA
AGGCGTAATA TCAGGAAAAA TCGAGGTCAG TGAACTACCT AGAACTGGCG
```

09/701 626

```
AGATAATTTC ATTCTCCTTT GCGCCAAACA AGTCTAAATT CCCGGCAGAG
CCAAGATTCA ACCCGTTGCT TAAAGTTGAG AGAGTGATTC ATAGCGTAAA
TGGTCAGAGT CCAGCTCTTC AGTTAGAGAA TCTGATGCTA CCAAACAGAG
AAAGTGTCGC TGAAGTCACT GCTTTCCTAG AGCAAGGCTT TGGCCTATTT
TTCAGCCCAA CCGGTGAGTA ATCCTCTAAC AATGCGCTCA AATCGCTCAC
TACGTTCGCT GGGACCGGCT AAAGCCGGCC CCTTAGCTTA ATCGTTAGAG
GTCAGCACAT GGCAGTGCAG CAACTCGGGC CAACCACAGT ATCCGTAACC
GAATTTGCAT GGGACGGAAG CGATCTTGGA AATACTGAGG CCAATGAATT
CTGGTCACAG CTCTCTGCTC AGCTTCAAAA AATAGCTATC TCTGAGTTTT
TAGCTGGCAA TCGCCCCAGC AGCATTCTTC GCAACGACCC ACGAAACATT
ATTGTTCTCT CATTTTCGGC GCCGCCAAAG TTCATTAAAA TCAACCACTG
GCTCTCTGCG TGTACACACA GAATTTCAAC ACGGAAATTA CTGCTACGAC
GGAAACGGCC TGTACTTACG AAAATTTAGA GTCTGGCGAC TTTCTTGCAT
TCGACACAGC GGCGTTGGTG CATGCCCTCT AACAATTGGT TCAAGTCGTT
CGCTTCGCTC ACTGCGGGAC CGGCTAAAGC CGGCCCCTTA ACCAAACGTT
AGGGCACCGC GCATGAGAAA TGAAGACGGA ACCTTTTGCA AAGACTGCCA
CCATCAACTT GATGAAACAC TAGCATCTAG CGCAAATTAC TCATGCCCCA
ACTGCGGCTC CACAAAAAAA TACATGAACA TGTCCATCAC TGATGGAATT
GGCCTATACG ACTCTTTGGG TGCCCAAGCT AAAGATCCAA GTTACCCGGC
AAAAGAAAAA TCAGATGGGA AACATTTGTT GGCTGGGAAC GCAGTCATAA
ACTGCAAAAA ATGGTTTACA AGACAAGAAC TATCGATCGA ACCAATGACG
CATACCAAGA AATAGTAGTC GACCTTAAAA CAGGGGGAAT AATTCATCAC
TGTGAAGAGC CACTTTCAGA GCAYTTKGGC CATGGCACCG CAAAACCAAA
GCCCTAACAA ATGGTTCAAG TCACTCGCTT CGCTCGTTCG GGACCGGCTA
AAGCCGGCCC CTTAACCAAA CGTTAGAGGT TACCTGTGAC AGATTCGCGC
CCGTTACTGA TCCCTGCCTC GCAATATGAT ACGAGCGTTC TTCTCGCCGA
ATGGCAATGG CTCACCCCCA AAACGGATAC GCCACTTTTT ATTTCCATAT
TCGGAGACTG GGTATTTGGC AACCCCAATG GAAGTTTGTG GGTTCTTTCA
CTCCTAAAAG GCACTTACGA GCAAGTAGCC GCAAACTCTA ACGAGTACAA
CACCCTCAAC AAATCGGCGG AGTGGATTGA TCAAACATTC ATCGCCAGTT
GGCAGTCTAT TGCCGCAGGC CATGGGTTAA TCCCAGAACC AAACCAATGC
CTCGGCTGGA AGGTTCACCC ATTATTAGGT GGAAGTTTTG AGCCAGCCAA
TCTCCAACTC TTCAACATGT CGGTGTATCA ATCGCTTATG GGTCAACTTC
ATCGACAGCT TAGCCAAAAA CAAACCCCGG CAAGTAAAAA ACCATGGTTC
CAGTTCTGGT AACCTCTAAC AACTGGTTCA AGTCGTTCGC TTCGCTCACT
GCGGGACCGG CTAAAGCCGG CCCCTTAACC AAACGTTAGG CGCAAGGGCA
ATATTGGTTA TTCAGCACCG AGCCAGGGAA CACAATCACC GCATCAGCGC
AGTGTTCCTG AATCGAATGG TCGCCTGACA GTAGAGGCCG TTATTTGTGG
CCAGCAAAGG AGTTGCTTTC AAAGAATGTA CACGTCACAA ATAACTTCCG
GGGCCAAAAC CGACACGCCG TGCGCACCGT CGGTCAAGCG CAGCGCTGGC
CTCACTTGCA GCGTACGGCT GGGCAATCTA ACAATTGGCT CAAGTCGTTC
GCTTCGCTCA CTCGGGACGT CCAATAGCTG CGCTATTGGC CGCCCCTTAG
CCAAACGTTA GGCCAACATA CTCAACGCAT GAAAACAAAA TATCACATAA
ATATAATTAT ATTTCTCGAA ATCATAATTC CTTTAGCACC AATAATTTGG
GCAATTTTCA CTCAGTCAAG CCCCGGCTTT GGCCCAACCC TTATATCAAT
GCTCATCCTG CACATCGTCG GACGAATAAT TAGCCGAAGC ATCCCTGCCA
GCTGTGACTC ATGTGCTGAA AAAATAAAAC CCAAAGGAAC CTCCGCAATC
TACTACAACT GTCAAAAGTG TGGATTTAAA TACTCAAAAA CACTTAACAG
CAGCAAAAAC TTCCATAACC ACTAACCAGA AAATCACTAA GGCGCCATCA
TGTTATAAGC GCCGTAAGCA CTAAAGACTT GTACAAGCCT AACAACTGGT
TCAAGTCGTT CGCTTCGCTC ACTGCGGGAC CGGCTAAAGC CGGCCCCTTA
ACCAAACGTT AGGGCACTCA ATGCATCGCT TCCTAGCCAC ATGCCTACTA
GCTACATCTA TTAAGGCATA CGCAGAACCT GAAAATAATA TCGACTGCAG
CAACGCATTC TCAACGCCGG ACATTGAACA TTGCGCATCA ATCTCTCTTG
AGAAAACAGA GAAAGAGCTA AATTTAGCAT ATCAAAAATT AGTCAAAGAC
CTTTCTCAGC CAAACAATGA ATACGAAAAT TTCACCGAGT ACAGGAAAAA
```

```
ACTTTTAACG GCTCAAAGAG CATGGATCGC GTTCAGGGAA GCAAACTGTG
CCACTCAGTA CGAAATGCAC AGATCTGGCA CTATTCGCAA CAGCATCTAT
CTAGCCTGCA AAGAAAAGCG TGCCAAGCAG CGAATAAAGG GAGCTTCAAA
ATTATGCTCC GTACTAGCCC TAACAAATGG TTCAAGTCGT TCCGCTTCGC
TCACTGCGGG ACCGGCTAAT GCCGGCCCCT TAACCAAACG TTAGGCCGAC
AATCGCAATT CCTAGGACTG CACGTGAACT GGATCCGCAA AATGTTTCGG
CGCACAGCAC TAGCGCCGCC CCAACATCGC GAGGACGAAG CTGTCAGTAC
AAGCCAAGAA GGAACGCCTC CCTTTCGTCA TTTGACAGTT GAGAATTCAT
GGGGAAGTTG AGGGCGGAGC TATTCCTTCA GTCACCACCC GAGAACATCC
TCAGAAGATC TGTTTCTCGT TTGGCGTGCC TAAGTTCGGA TGGTCAACGT
TCGAGATCCA TTTCGTCGGA AATGGCCACT TCATCTGCGG CATCTCTGAC
ACTCCAAATG ACTTCTACGG TGACTTGGCT ATCGCCCTGG CTGAGCAGAA
AAGTTCTTTT TCGGTAGCGG CGCACCTTGA GCCTGAGACC TTTGCCTTCT
ACATCGTTGA TTCGACAATG TACTTGTGCA AGTTCGATGA ATTCGACGAT
TATGAGTCCG CCGCCGAAAG CCACGAACAG TTGGTCTCCC ACAGCTTTAT
GTCCATTGAA GTATCTAGGG AGTACTTTCA GAAGTCTCTC AGGACCTTGG
CCGTCCAATG GCCGGATACG CCTTCAAGAG ACTGGGCGCA CCCATTTCCA
CGTGCGCAGA TTGAAGGCTG ACTGCCTAAC TATTCGCTCA AGCGGGCAGC
GTTAGGCGCC CTCATTCGGA GTCACGCTAT GGCAACCCGA GAAGAAACAG
AAGTAGCCAT TGCTGCTCTT CGCAGCGAAC TCAATGGCAA CGAATCGGAA
TACAGCTTTC ACATTCCCGG TTGGGCGCCA GAAACATCAG TCATGGGATT
TCGCTGGATG CAAAGCCAAC TGTGGGAAGG CTTCTACGTA AGCTATCGCG
TAGAGCACTC GGCCAAGCGC GTCGAATTCA AGTGCTGGGA GTACGGCGAG
CCCGAGCCGT CTTGGCTGCA AGTTGGCTAG GGGGCCGGCA AGATGCAATC
GCGGCGAGCG CCTAACACTG CAGTCAACCG GACACCAAAC TGTACGCAGT
TTGGTTCCCT CCGCTGCGCT CCGGTGCCGG TTACTTTCAA CGTTAGGCAA
CTCAGATGAG TGCTCCAGAC GCAGAACTTC TCGCATTGTT AGCCTACCGA
ATGGAAGCTA TTTCCATTGG GCATTTGGCA TTACGCCATC ACATGACGTG
GGACGAAACA CCTTCAATGG AGGTGTACTT CAATGGCATA CAAGTACTCG
AGGGAAAGGC CACGGGTTTC ACTAATGCAG CCATTGAGTC CGCAATTATT
CATTGCAGGG CAATCCTTGG AGTTTGTTGG GCTGCAGTCC TCCAGACACT
CTTCCACAGA AATTGCAGAG CGCACTCGAC GCAACAATCC CGATGACTAT
GGCATTGAAA GCTTCAATGG CTTATCAATG CTAACCAAGG AAAAAGCACT
AGCCTACTAC TCTGGCGAGC TGCCAGAAGC GGAAGTTGCT CTAGCGCTCA
TATTCCACTC AGCGAACAAA GGGCTTGCAC ACACTACAGT GTCCTTTACG
CGTGACAGTG GCGACGCCCA CCTGATGGAA ATTGCATTTC GCATCGTACC
AATCCTGCTT GTAAATGGCT TCTACGCTCC ACTGGAAATC ACGCCACCAA
AATATGAACT GATTTCACGC CCAAGAGTCG CCATAACAAA TGGTTCAAGT
```

09/701626

Figure 3D-1
SEQ ID NO:4

```
GCCGGCCTGC AACAGAGCTT CAAGGCCGCC GGTGTCGGCA TTCGGCGACC
GCGTGCCGCC TTTATGCCAA CACCTGGCAG ACGGTGGTCG GCATGTTCGC
CACCGCCAGC TTGGCGCCAA CTGGTCGTCC TGCTCGCCGG ACTTCGGCAC
CCAGGGCGTG ATCGACCGTT TCGGCCAGAT CGAACCCAAG GTGCTGATCG
CCGCCGCCGG CTACCGCTAC GCCGGCAAGA ACCTCGATCT GACCGCCAAG
CTCAACGAAA TCCTCGAACG CCTGCCCTCG CTGCAGCAAC TGGTGGTGGT
GCCCTACTCC AACCCGACAG CCGGGGCGGG CGACTTCGC AGCGCCGCCC
GTGTCAGCCT GTGGCAGGAC TTCTACCAGG CCGGCGGTGA ACCGAAGTTC
ACCCCGGTGT CCTTCGAGCA GCCGCTGTAC ATCCTCTATT CCAGCGGCAC
CACGGGCGTG CCCAAGTGCA TCGTCCACGG TGTCGGTGGC ACCCTGCTGC
AACACGTCAA GGAACTGGGC CTGCATACGG ACCTGACGGC CGACGACACG
CTGTTCTACT ACACCACCTG CGGCTGGATG ATGTGGAACT GGCTGGTCTC
AGGGTTAGCC TTGGGCGCCA GCCTGGTGCT GTTCGACGGC TCGCCGTTCC
ACCCAGGTGC CGAGCGCCTG ATCGACCTGA TCGACGCCGA GAACATCAGC
CTCTTCGGTA CCAGCGCCAA GTTCATCGCC GCCCTGGAAA AGGCCGGCGC
CAAGCCGCGC GAGACGCACA GGCTGCGCCG CCTGAAGGCC ATCCTCTCCA
CCGGCTCGCC GCTGGCCCAC GAGAGCTTCG AGTACGTCTA CCGCGATATC
AAAAGCGACG TCTGCCTGTC CTCCATCTCC GGCGGCACCG ACATCGTCTC
CTGCTTCGCC CTCGGCAACC CGACCCTGCC CGTGTGGCGC GGCGAGCTGC
AGTGCAAGGG CCTGGGCATG GATGTGCAGG TGTGGAACGA GGCCGGCCAG
CCAGTCATCG CTGAGAAAGG CGAGCTGGTC TGCGCCCGCC ACTTCCCGTC
GATGCCGGTC GGCTTCTGGA AGGACGCCGA TGGCGAGAAA TTCCGTAGCG
CCTACTTCGA CACCTTCCCC GGCGTCTGGG CCCACGGCGA CTATGCCGAG
ATCACCGAAC ACGATGGCCT GGTGATCCAC GGCCGCTCCG ACGCCGTGCT
CAACCCCGGC GGCGTGCGCA TCGGCACTGC CGAGATCTAC CGCCAGGTGG
AGAAGGTCGA GCAGGTGCTG GAGTCCATCG CCATCGGCCA GGACTGGGAA
GGCGACGTGC GCGTGGTGCT GTTCGTGCGC CTGCGTGACG GCGTGGCGCT
GAGCGACGAA CTGCAGGCAC AGATCCGCCA GGTGATCCGC GCCAACACCA
CGCCGCGCCA TGTGCCGGCC AAAATCATCG CCGTCGCCGA CATCCCGCGC
ACCATCAGCG GCAAGATCGT CGAGCTTGCC GTGCGCAACG TGGTGCACGG
CAAGCCAGTG AAGAACACCG ATGCCCTGGC CAACCCGCAA GCACTTGAGC
TGTATCGCGA TCTGCCGCAA CTGCAGTCAT GAGCCGGTAA GCGACACCGT
AAGGGCAATG GACTGCCACT CCAGCATCTA TAGTGGATGC GCATAACCCG
GACAGGATGT TGCTGATGCA GGTGGTTTTC TGCCTGCTGG TTGCCTTGCT
GTATGTCGGC GGCGTGGCCG CTGACGAACC ACTGGCCTTG CATATGCCGG
ACGCCCCGCC GCTGACCCTG TACCACGACG AGCGCGGCCA CGGCATGGTC
GGCGACATCA CGCTCGCGGC CATCACTCTC AGCGGCCGAA CGGCACGCAT
CGTCGACGAG CCCTGGGCCA GAGCCCAGGT GAACGTCGCC AGCGGCCAGA
ATCAACTGAT CATCCCGCTG TCGCGTACCC CGGAGCGTGA GCAACGCTAC
ACCTGGATCG TCCCGATCAT GCCGCTGGAG CGCGCCTTCT TCAGCCTCGA
CAAACCTGTC AGCAGCTTCG CGCAGGCACG CCAGCGCTAC CGGCGTATCT
GCGTCGGGCT CGGCACCGCT CAAGTGGAAA TCCTGCGGCG CGAGGGTTTC
GCCGACGAGC AGATCATCCA GCTCAAACTG GGCGAAAACC CGGCCATCCT
GCTCGAACGC GGGCGTCTCG ATGCCTGGTT CACCGGGATT CCGGAGGCGC
TGTACATTTG GCACAAATCT GCGGAACAGC GCCGCAAGCT TTATCAGAGC
CCGGTCCTGG CCAGCACCGA CCTGTACCTG GCCTGCTCCA GGATCTGCGC
CCCGCAGATC GTCGAGCAAC TGCGGGCCGC CGTGCTGCAA CTGGAGGCCA
GCGGCGTCAG CCCGCGCCTG CGCCAGGCCT ATCTACCCGA GCTCGATCGA
CGGTGAGCAC CCAGCCCACG CTGCCAGGCC TATAGAACAG ATCATCCACG
CCGCAGGCCT GCGCCGGCTT GAGCATAACG CCAGCACTGG GGCAGACTGC
GACTACCCCG CACAACCCCG GTGACAATAT GGACGTTGTC AAAACCCTCA
AGCCCGGCAA ACCCGGCACC AAGCGCTTGC AAGAACGCTA CGGCGAGCAA
```

```
CTCGTCGCCG TCCGCTACCG CCTCGACCGC AAAACCAACA CCCACTACAC
CACGGTCGAA CTCATCGTCG AACAAAAGTA CGCCCTGTAC AAAACCCCGC
CACCCGCTCC CACACCTCCG GTAGCCCTGC GCATCTTCCG CCACGAAAAC
GACCTCCAGC GACTGATCAG AAGCGCCGGC GGCAAGTGGG ACCGTGAGAA
TCAGGTGTGG CTGATCGAGC GAAGCGAGGC CGAGAGGCTG GGGCTGGCGG
AACGGATCAT CTGGACATAA TGGCTATATG TGGACATCAA GATGCCTAGT
AATAGCCACA AACACCCAGC ATCGGACACT ATGCCTACCC CTAGGCATGT
CGTATAAACA CTAGTTATAC AAATATCATA TGAACGACGC GACCCTAAAG
CTAGTTAATC AAAGACAGCT CGTATCGGTA ATGAATAAAA CGAAGTGGAC
TGAGCTGTGC AATTCATTTG ACTGCGAGAA TAAAGCATCT CCGAATGTTC
GCTATAAATT AATTTACAGT GAACAAGAAT TCGGTTTTTC AAAAATATGG
TGGAATCAGC TTTTGCATGA GTGCGAAGCA ATCGAATGGA TTGATTTCAA
ACTAGTATTG CGAGAACACC GTGGCAATCT ATTGCCAGAC AAAGAAATTG
ATATAAGCAA ACAAATTAAG GAAGCACTAC AGGCGCATGA CATCCCTTAC
TCTGTTGAAG GAGAAAATCT TAGGGTTTGG GGCTATATTA GCGCAGAAAA
GAGTCCAGTA TTCGTATAAC AATTGGTTCA AGTCACTCGC TTCGCTCGCT
CGGGACCGGC TAAAGCCGGC CCCTTAACCA AACGTTAGAT GCTTATGAAA
AAGACAGTTC TCATACTCGT CCCAGCATTA CTACTCTCAG GATGTGGCGA
CCCTGAATTT CACTACCAAA ATGGTGACGA ATCAAAAAAT ATAACGCTAC
GCATCCCTAA GAATTACATA AATTATTTCC CTGGCGTGAA GTACGAAAAA
GACGGACCTG TCGTCATCAG ATTTTCATAT CCACAATTGG AGCCACTGAC
AAAAGCCCTA CCAGAAGAGC AAAAAGTAAC TGTCAGCATT AGTCATTTAT
CCAGCCTGGA ACTCACCACC CAAGAAACCA GAAACCCCTA CTGCGAAACA
GATAAAAAGT GGAAACTCCT ACAGGCGGCG GGCATTCACG GAGAGTTCTA
TAAATTCATC GGAAAATCTC CCGGCAGCGC CAGTGCAGAT ATAACTTATA
AGCCCATCAA AAAAACACTT GGCCTTTACT GCATTACATG CGTGGAAAAT
GCAAATTGTG AAATTCACGC AGTATCTAGC CAAGGAATAA GCTATTCCGC
ATTTTATACA GAAGACTTAA TGCCAGATAA GTGGCACTCT ATTTACATGG
CAGTCGACAA AATCCTTAGC AAATTTACAG CATCGTCGAA AGGCATCTAA
CAATTGGTTC AAGTCGCTCG CTTCGCTCAC TCGGGACCGG CTAAAGCCGG
CCCCTTAACC AAGCGTTATG CAAGCAGTCA CCCATGAGGA AAGCACCCAT
ATGGAGCCAG TATGAAATTG AGCGACATAA GAGCTCTAAT CATTGAGTCG
CCAGGATGGC GAACAGTATT TGCATTTATT GTCCCACTAA TCGCAGGGAT
TCTGTCGGGA ATATTCGTAT CAGAAATAAC GCATAGCTCC GAAATTGTTT
GGAAGGAATT TTATAAAGCA AAAAGCTTCT ACGGGCTATT GGCTTTGAGC
TTGTGCATGT ATTTTTACAA TAAAGCCATT TATCTACATG AAAGAGAAAT
TTCTCGCTTC CTAGACGCAG ATTACTGCAC CGCTTACATG AGAAGCAAAT
GCCTGCCAGA GGCTGCAGAG CGATACAAAA AGCTTATACG CTCTGGCGAC
GGCGGCGAAT TGAAGCAAGC AATGGATGAA CTGAAGAAGG TGCTCAAATG
AAAGTACTGG CCAGCCAGA TTTTAATGCA AAAGTGCCGG CACTAAACAC
AGAAACCATT AGTAGCCTTT CTGCATTCAT ATCAAGCGCA GAGCAATATG
AAAAAAATGA CTTCATATTG AAGAATGTAA ACTCAATGTC TCTTCTTGAT
GGCGATATAT ATAGCGCAAA AATCAACTCA AGCAGACTAT ACTTCACCAT
CGGAGCTGAT GAGCAAGGCG ACTACTTGCT GCTATTAGAT ATAGCTGCCT
TGCAAACCGC ACCATCTGTC AAAAGTAGTG CTTTCTTCAC AACAAACAAC
CCAAAAACCA ACAGCTCACT CAATCCGAAG CTCAACTCTG CAATCAACCC
AAAGCTAAAC TCAGCAATAA ACCCAAAATT AAATTCGGCC ATAAATCCGA
AGCTGAACTC AGCAATAAAC CCAAAGCTAA ACTCGGCCAT CAACCCGAAG
CTGAATTCAG CAATAAATCC GAAACTAAAC TCAGCAATAA ATCCAAAGCT
AAACTCAGCA ATAAACCCAA AACTAAACTC CTCACTAAAC CCAAGGCTCA
ATCGAAGCTA TGGCGGCCCG TATTTGTACG ATGCGAACCT TAATCAAGAA
GCGTACTCAG TTAGAGCCAA TAACAAAATC GAAATCCTGT TCAATTCGGG
CGGAGATTTT TATGGCTTTC TTGTAAGCGC TAACGACCGA GTGAAGATTG
AGTTCGATAC AGGAAATACC TGGACAGGTT ATTACGTTAA AGCCAATGAA
AAAGTTTGGC TTAGATATTC GCTTAACAAC GAATGGTTAG GGCTACTTGT
CTAGCCCGCA TAACAAGTCG CTCAAATCGC TCACTTCGTT CGCTGGGACG
```

PCT/US99/13295

```
GGCTAAAGCC CGCCCCTTAG CTTATCGTTA GGCAAAAAAA TAGCAGGCAG
GCTCAGTAAT ATGAAGTTCG ATAGAATAGC TCGTGAAGCG TTTGGCTCAG
TGCTTGGTCC ACTGGGGTTC AGCTGTAGTG AGTCGAAGGC ATGCACCTTC
TATAAAAAAG TCGGCACTGA GCTCTATCAT TTTGTCATGC CAGATCAATT
AAGCGGCCAG GAAAAGTATG ATATTAAAGT TTTTTTCCAC TCGCCGCTCT
TAGAGCCAAC CGCATGGAAT GACAAGTTTC CGGACACCTT GGGGATTCCC
ACAGATAGCT GGAGTTATCT TTCTAGCCGT ACTGGCGTTG GTCCACGACA
AGAGCTGTTT TGGTGTCGAA CAGAAGAAGG ATTTATGCGT AACTTTGAAT
CAAAGGTAAA GCCCGCACTA CTTCAATTTG TAGCCCCATA TTTTGATTCT
ATACAGACAT TGGAAGAGGC TATTCCACTA ATCAAGAGCA GGCACTATGT
GGCAGTGGCG TCTACGCTAA ATGCTAACTA AGCAATGCCA AGGTCTTCCA
CCGGCACCTC CGTATCGGCC TTGACAGATA GCAGCAATGA GTTTCCAGCA
AAAACCAATG CGCCGCTTGC AAGGCTGTTT CGGGTTAGCC ACAGTGCGGT
ATTCATTACC TGCGTCCGAC TCGATACCAA TTGCCTAACA ACTGGTTCAA
ATCGCTCGCT CCGCTCGCTG GGACCGGCGA AGCCGGCCCC TTAACCAAAC
GTTAGGCTAC ATATGAGAAT CAGCGCAGAC CAGCTTGCTC AAGAATCACT
GACTGAGTTC GGCGTGCTGG CGGCTAAGCT TCTGGCAACG CGAGAGCTTA
GCCAGTTGTC CGAGAAGTTT GGGTATGCAC TGGCCTTCGG AAGGGAACCG
GCGGCTGCCA TAGCTGAGGA CCTTGCTAGG TGCTTGTGCG GACAAAATGC
TTCGCCGGCA TCTGAATACC CCAAAATCAC CGTTAAGTAT TTCAAGGAAA
ACGAAAGTAG TCTGTTGGCA CTCGTAGAGT GTTATGTACA AATGACCGCA
AGCGCAAACA TTCTTTTAGA GCTGGTTGCC GCACGAAATG GAGAGGCAAT
AAATCTGTAT CTAGAAGGCT TGAGTGTTGT AGCCTAACAA TGCGCTCAAA
GCGCTCACTT CGTTCGCTGG GACCGGCGAA GCCGGCCCCT TAGCTTAATC
GTTAGGTGCC TCAGGAGGGA TCATGTCTTC CACAGAAAAC AATAGTGATG
ACTGGCGAGA AATTCGAGCA AGAGCGGACT CTATCGCTAA TGCCATTTTC
CTCATTTCTG GCGGGGCACT TTCACTTTCA ATCTCAGTCA TCCTCAGCAA
CAAAAGTGCC GGGTACATCA CTGCACAAGT GGCATGTATT GCGTCCCTCG
CTTGGTACTG CTTGCTGGCG TCACTGATTC TCTTTCTTGC TCTTAAGGGG
CATATGATTC TTCAGGCATA CCTCCTACAA TTTCGCCCAA ATTACGTCAA
TAAACATCTT AGATTTCTTA ATGGTATAAG CTGGGCCATT GGATTAACCG
GGTTTATTTC CTTCATTGCA GGCATGTTTC TTATGGTTCG TACCGCAATA
CTTGCCGTCG GCACCTAACA ATGCGCTCAA CTGTCGCTCA CTTCGTTCGC
TGGACAGTCA AAAGCTGCGC TTTTGCCTGC CCGTTAGCTT AATCGTTAGC
GGTCATAAGT ATGCAGATTA ATTTCTATAT GGCAGATGAA GATCGAAGAG
CGTTCCACGA ATACCTATAT TCTCGTGGCG CATACCTCGT TCCGGAGCGT
TGGCCAACCA GAGATATTCC CATAGTCCAG GCGGCCTCCG AAGAGGCAAG
TGAGTGCAAA GACTTCAAGA TTTTCAAGTC TGACCTCTTC CCTCAGTCCG
AATTTCAGAA CAGGGCTTGG ATAACGTGGC ATGAGCCAAC GAAAAGGTTC
TACGTTCATG GGCCTGGAAT TCAGTATCTT GTATCGTTCA CTGATGCAAA
TGGAATCCAT CGTGGCCGCC TCTATATGGG CCTTGTTTCT CCGCGTAGCT
TTGTTGAGCC CCACGGGCAA TCAGTTGATT GCTACGCTGA AAACGAGAAA
AAGTACAAGG CGTTAGAGAA TTTCCATAAG AGTTGCGCGC GCTATATACG
CAATCACTAC CGCAAAGATG AGGGTGGTTT CTACCATGGC AAAGCAAGCG
ATATGGCTGT TCAAAACTAC GGCGTCTCAA AGACGCAGCT GTGACCGCTA
ACAATTCGCT GCAGGCGCGA CGGCCCTGAC GGGCCGCGGC CTGAGCTCAA
ACGTTATAAC CTACAAGGAA GACCAAAGTA TGCGCCACCT AGCAATAGCC
CTCTTAATAA TGTTCTCTAC TCAAGTTCTC GCCGACGGCA AGAGCGAAAA
GATAGAGAGC CTAATGAAGG CACTTGGACT AGTAGACACA TGGACACAAC
AAATTGAACA AGGAAAAATT TACAACAGAA AGATCAGCTC TCAAATGCTG
GATCAAATTT TATCCAGCT GAATCCAAAT GAAGAGTTTC AGCAAAAATT
CAAAAAGGCT TCAGATAATT TCATAACAAA AACAGAATCT CCATGGTCTC
CAGAAAAAAT TGTAGAGGTT TGGGCTAGTT ACTATGGCCC AGAATTCACA
GAGGACGAGC TTGACCAATT AATTGCATTT TATACTTCCC CTCTTGGCCA
AAAAGACATC CGTGTTACTC GCAGTTCAAT GGAAAAATTC TCGAAATACT
TCCAAGAGGC CGGGCAACCA ATACTAGAAA AGGCCACCGC AGAGTACATT
```
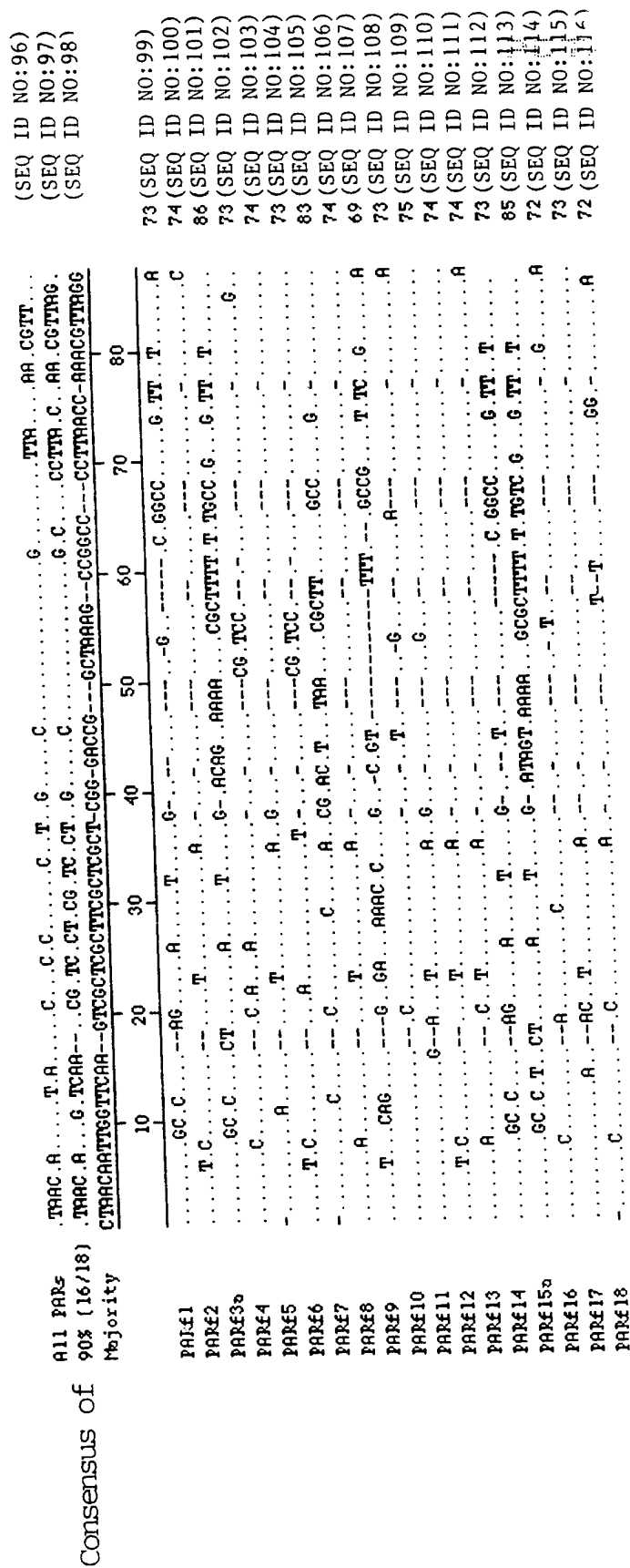
```
CAGGAAATGA AGCTCATCGC CAAGGAATGT AACTGTACCA AGTAGCTTAT
AACAATTGGT TCAAGTCGTT CGCTTCGCTC ACTGCGGGAC CGGCTAAAGC
CGGCCCCTTA ACCAAACGTT AGGCACTGCT ATGGCCTTGG TCGAGTACGA
ACTGATCATC AATGCGCCCC AGACGGCTGT CTATGCCGCA TCTCAGGACT
ATTCAGTTAG GTACCAGTGG GACCCCTTCC CTGAAAAAAT TGAACTCCTA
GGTGGTGCAA CCGAGGTAGG AATTGGGGTT AAGACACTTG TAGTCGCCAA
GTCTGGCTTA ACAATGGAAG TCGAGTTTGT TCAGGTTGCT CCTCCTACAA
CGGCAGCCAT AGTCATGACC AAAGGCCCAG CATTCATCAA GAGCTTTGGT
GGTAGCTGGG TTTTCAAGCC CATCACCGCA AACTCTACAA AGGCAAAATT
TCGCTACTCC ATAAAAACCA AGAAATGGGC AATACCCATA ATCTCAGAAT
ACGTAGCAAG TCTTTATTTC AGAAGAGCAG TTAAGGCCAG GCTTGCCGGT
CTTAAAAAAT ACTGCGAGCA AGGCGCCTAA CAAATGGTTC AAGTCGCTCG
CTTCGCTCAT TCGGGACCGG CTAACGCCGG CCCCTTAGCT TAATCGTTAG
GCTGGCCGAA GATATGAGTT ACAAGAGATG GATTTGTGTC CACTGCGATA
CAGCCAACAC CACAGCAACA GATATTTGTT CAAAATGTCA CAGATCCAGC
TATGAAGAGC CGGCAATAGC TGAAACTCCA ATAGCTAATT CTTACCAAGG
CATACAGCTG TTAGGCTCTT GGCTTTTTAT CCCACTAACC CCATCCATTA
TGGTAATTGC AATAAGGGAT GAAGTCTGGT GGTTCGTCCC ATTTGGGATC
GCAGTTATTG CGCTCACAAT ACTAAGTGAA AAATCTAAAT TCTTAATTTC
CAATACTACT TGGTTCAAAA ATATAGCTTT ATTTTATACC CCAGCAGCGG
GTGTGCTTTT CCCTCTTAGC GTTTTTCTCG GAAAGAATTG GGCGGCCGCA
TTTATGGCAA TGCACGTGGT TGTTCACCTA CATGCTGCAT TTAACATGCA
CGCACACTGC CAAAACCACA AGCCAAGAAA TGAAAATTAA CGATAAGTAC
TCTAGGCCTA ACAATAGGTT CAAGTCGCTC GCTTCGCTCA CTTGGGACCG
GCTAAAGCCG GCCCCTTAAC CAAACGTTAG GTTCCATCAT GTGCTACATG
GCCATAGTAA GTACAACATC CGAATCTGAC CTAACGGCAC TAAATACGCC
CCTAGCGCAG TTCTCAAGGA ACGTAAACGC CATACCAGAA GCGGCCCTCC
TGCGTTACCC AAATAAATGG TTCCTTGGCT CAAAAGACGG TTGCAGTTGT
GCATTCAGGC ATCTTGATCA AAATGCTACA GATCTTGGTT TCTCAGAGCC
CGTGGACTGG TGGGAAGAAG ACCAAGAAGA TATAGATGCT ACGCTTCAAG
TAGTAGAAGC ATTCCATACG ATATTGCGCG ACGGCCATAA ACTTGACTGC
ATAGATGCTT GGGCCAACGA TAGCAAGGAA CCTAAAAATC TTGCAGGTGA
TCTTGTGGTT GACCTAAATA AAGTTGGCGC CAAGAGTTTC CGTTTTTTTG
AAGGCTATCG CTTTGAGCTC GAAGCCAGAA CCTAACTAGT GGTTCAAGCC
GCTCGCTTCG CTCACTCGGG ACCGGCTAAA GCCGGCCCCT TAACCAAACG
TTAGGCCCTA TATGCAGTAC TCAATTGCTG ATACTGAGAG TTTTCATCCC
GTCATGGATG CTGAGATCAA GGGGCACTGC GAATTCCCCG TTGACCTAAT
TTTTGTCCCG GACATTGCAG AGTGGGCTGC TTCCAGATGC GGAAATCTCA
TCGGCAATCC TGTCGCAATG GCAGTTAGAG ACGGAGCCAC TAAGGGGCA
GGCATTCTCA TTAGGCAATC AATAGACGAA TCGCAAGTCG ATAGCATCCT
ATCTCGAATG GAGTTCGGCG GCTTTGATCG TGCGCGGTCC ATACTGTCCT
CACCCGAGAA ATTTATGCGG CATCTAGTCC TCCATGAGCT CGCGCACTTG
ATTAACAATT GGGGGCAAGA CCGAGAAGAC GACTGTGATG AATGGGCTTT
TAAGCGTCTT GGTGCTAGGG CCTAACAATG CGTATATGGA CTCTCCCCAC
AAGTAGTGGG CAAAGCTTTT GCTCTGCTCC TTTCGTCGGT GCGGTTCCAT
GCGTATATCC GGCCTTTCGC GAGCTCAGTG CTCTGGCCAT TCTGTAGTTC
GCGCAGCGGG GGCCCAGCGC TCAAGCGATC TCCAGGATCA GCATCGTCAC
AGGCCTGCTG TGCAGCCGGC CGAAGGAAGA CCTTCCCTCT GCTCTCCCTC
TCCCCCATCG CTCCTAATGG GCGCGTCGGC AGCCCAGGCT GCCCCAGCCT
ACAAGGAGCA AGAACATGCC AAACCTGACG TTGCATCGCG GTGAGAAGAA
CAACTGGTGG CCAGCACCCG AGGTGCGGCT GATCTGCGGG ATGACCCTGT
ATGGCCCCTG GCAGCCGACG GCGGTCAGCT CGCTCTGGAA CACACTCAAG
AACGAGGTGA AAGGCGCCCT GGGCACCTCG CTGGAGAAGA AGGTCGCCGC
CTACGCCCAA TACCTGCGGG CGACCGGTCG CCCCTTTGCA CTGGCGACGG
CCTGGACCGA AGTCGGCTCC TTCACGTCCG ATTACAACTA CGTGATCCGG
ATTCCCAATG CACATTTCTT CTACTGGGGT GGCACCAAGG ATGCCCCGGA
```

```
CCTCGGCGCT GCCGCCGCGT GGACGACGCC CGAGCAGGTG ACAGCCGACT
TCATCGTGCT CAACGCTCCC ACCGTGGCGG CCTCAACCAT CCTGGGCTTC
GGGCACCACA CCGGCACGCG GGAGATCACC TTCTTCCACG ACCTGCCGAT
CGGCCTGATC GAGTCCTGTA ACGGCAGGCC GATCACCGAC TATGCGATCA
AGAGCAAGTC CGACCTGAGC TTCGACGAGA GGATCAAGTA CGCCAAATAC
CTGCGCTAGC CGCGTACCCC GCGTCCGAGA GGCTTAGAAG CTAGGGCGGC
CGGGGTCTTC CGGGGGGGTG TCTTCCTCGA TTTCCTCAAG CTTGAGTTCC
ATCGCCCAGT TGGCCGGTGC CGCCGTGGGC GCGGCAACGG GTGCGGGCGC
CGGGGCGGCA GCCTGCGGGG CGGTGGGGTT GTCCTTGTAC AGCTTGAGCT
TGAGGCGCAC GTTGTTGGCC GAGTCGGCGT TCTTCACTGC CTCCTCCTCG
TCGATGACGC CTTCATGAAC GAGGTCGATC AGCGCCTGGT CGAAGGTCTG
CATGCCGAGG TTCTTCGACT TCTCCATGAT CTCCTTGAGC TCGGAGAACT
CGTTGCGCTT GATCAGGTCG CGTACGGTCG GCGTGCCGAG CATCACCTCT
ACGGCGGCGC GGCGCTTGCC ATCGACGGTC TTGACCAGGC GTTGGGAGAC
GAAGGCGCGC AGGTTGTTGC CGAGGTCGTT GAGCAGCTGC GGGCGGCGCT
CTTCGGGGAA GAAGTTGATG ATGCGATCCA GCGCCTGGTT GGCGTTGTTG
GCATGCAGGG TGGAAATGGC CAGGTGACCG GTGTCGGCGA AGGCCAGGGC
GTGCTCCATG GTTTCGCGGT CGCGGATCTC GCCGATCAGG ATTACATCCG
GCGCCTGGCG CAGAGTGTTC TTCAGCGCGG CGTGGAAGCT GCGGGTGTCC
ACGCCGACTT CGCGCTGGTT GATGATCGAC TTCTTGTGCC GGTGCACGTA
CTCCACCGGG TCCTCGATGG TGATGATGTG GCCGCCGCTG TTGCGGTTGC
GGTAGTCGAT CAGCGCCGCC AGGGAGGTCG ACTTGCCGGA GCCGGTACCG
CCGACGAACA GCACCAGACC GCGCTTCTCC ATCACCGTCT GCAGCAGCAC
CTCGGGCAGC TTGAGGTCCT CGAACTTGGG GATGTCCATC TTGATGTTGC
GCGCGACGAT GGATACCTCG TTGCGCTGCT TGAAGATGTT GATGCGGAAG
CGACCGACAT TGGGCACCGA GATGGCCAGG TTCATCTCCA GCTCCTTCTC
GAACTCGGCG CGCTGCTCGG CGTCCATCAC GCTATTGGCG ATGGCGGCGA
CGTCACCCGG CTTGAGCGGC TCCTGGCTGA GCGGCTTGAG CACGCCATTG
AACTTGGCGC AGGGCGGCGC CCCGGTGGAC AGGTAGAGGT CGGATCCGTC
CTGGCTGGAC AGGATTTTCA GCATCTGGGA AAGGTCCATC GCACGCGCTT
CCATTTGGGT GGAGTTAACA AGGTAGGCCA GCTTTGCCCG GCCGATCAGC
CTGAAAAATG GCGCCATTCT GATGGCGCAA CGAATGCTGG CACAATAGCG
CCATCGCAAA ATGAGGACCC CGTCATGCCC AAAGCCATGG CCCGCCACAT
CCTGGTGAAA ACCGAAGCCG AAGCCGCCGC CCTGAAGAAA CGTATCGCCG
CCGGCGAGGC CTTCGATGTG CTGGCAAAGA AGTACTCCAC CTGCCCCTCC
GGCAAGAAAG GAGGCGACCT GGGCGAGGTG CGCCCGGGGC AGATGGTGCG
CGCCGTGGAC CAGGTGATCT TCAAGAAGCC CTTGCGCGAA GTGCACGGCC
CGGTGAAGAC CCAGTTCGGC TATCACCTGA TCCAGGTGTT CTACCGCGAG
TGATCCAGCG GCTTAGCCGG CCCAGCCGAG GGTAATGGCG GCCAGCACCA
GGTAACGGCC GGTCTTGGCC AGGGTCACCA GCAGCAGGAA GCTCCACCAG
GGCTCGCGCA TCACCCCAGC CATCAGCGTC AGCGGGTCGC CGATCACCGG
CGCCCAGCTC AGCAACAGCG ACCAGCGGCC ATAGCGCCGA TAGGTGTGTT
TGGCCTGCTC CAGGCGTTGC GCGCTCACCG GGAACCAGCG GCGCTCATGA
AAGCGCTCGA TGCCACGGCC CAGCGCCGCA TTTCAACACC GAGCCCAGCA
CATTGCCCGA TACTGGCCAC CGCCAGCAGC ACGAACACAG GCTGGGCGCC
ACCCAGCAAC AGGCCGACCA GCAGCGCCTC CGACTTGCAG GGGCAAGCAG
GCTGGCGGCA CCGAAGGCAG AAAGAAACAG GCCGAAGTAG ACCGAAAAGT
CGAACACAGG TGCCATCCGG CAAAAAGTCG GG
```

Fig. 3E

Alignment of *Pseudomonas alcaligenes* repeat
(PAR) elements from Contig 1

Consensus of	All PARs	(SEQ ID NO:96)
	90% (16/18)	(SEQ ID NO:97)
	Majority	(SEQ ID NO:98)

```
                    10        20        30        40        50        60        70        80
All PARs   .TAAC.A....T.A....C..C.C.....C.T.G.....C...........G...........TTA....AA.CGTT...        (SEQ ID NO:96)
90% (16/18) .TAAC.A...G.TCAA...CG.TC.CT.CG.TC.CT.G.....C.......G.C..CCTTA.C.AA.CGTTAG.        (SEQ ID NO:97)
Majority   CTAACAATTGGTTCAA--GTCGCTCGGCTTCGCTCGCT-CGG-GACCG---GCTAAAG--CGGGCC---CCTTACC-AAACGTTAGG  (SEQ ID NO:98)

PARf1      ..........GC.C.....-AG.....A......T......G.......-G.........C.GGCC.......G.TT.T.......A  73 (SEQ ID NO:99)
PARf2      ..........TT.C........T......A......T..........-.............-----...........C  74 (SEQ ID NO:100)
PARf3a     ..........GC.C....CT....A....T..G-.ACAG.AAAA...CGGCTTTT.T.TGCC.G...G.TT..T.  86 (SEQ ID NO:101)
PARf4      ..........C..A....-C.A..A.............-------...CG.TCC...-----......G.  73 (SEQ ID NO:102)
PARf5      .-...A......T......A.G........-----......-----..........  74 (SEQ ID NO:103)
PARf6      ..........TT.C.....T..........T-......-CG.TCC...........-----...  73 (SEQ ID NO:104)
PARf7      .-...........C......A..........C...........T-......CG.AC.T.TAA...CGCTT...GCC.......G.-.  83 (SEQ ID NO:105)
PARf8      ..-........A......T............G.-C.GT......-----..........TTT.-.GCCG....T.TC..G.....A  74 (SEQ ID NO:106)
PARf9      ..........T..CAG...-G.GA..AAAC.C...G..-C.GT....-----.....A----..........A  69 (SEQ ID NO:107)
PARf10     ..........C.......-C.......A.G.....-----.....-G..........-----..........  73 (SEQ ID NO:108)
PARf11     ..........G-.A....T......A.G.....-----.....-G.........A----..........  75 (SEQ ID NO:109)
PARf12     ..........TT.C.....T......A..........-----......-----..........  74 (SEQ ID NO:110)
PARf13     ..........A...-..C..T.......G.-......G....---.T......-----...........C.GGCC.....G.TT..T.  74 (SEQ ID NO:111)
PARf14     ..........GC.C.....-AG...A....T......G.-ATAGT.AAAA...GCGCTTTT.T.TGTC.G...G.TT..T.  73 (SEQ ID NO:112)
PARf15a    ..........GC.C.T.CT...........A......T.......-----.........T......G.  85 (SEQ ID NO:113)
PARf16     ..........C....-.A..AC.T.....C......A...........-----..........  72 (SEQ ID NO:114)
PARf17     ..........C.......-.C.........A......-----.....-----.........TL-T........GG.-.....A  73 (SEQ ID NO:115)
PARf18     ..-.........C.................-----......-----..........  72 (SEQ ID NO:116)
```

09/701626

DOTPLOT of: superintegronI.pnt   Density: 16075.00   May 25, 1999 16:22

COMPARE   Window: 21   Stringency: 14   Points: 52,932

superintegronI.seq ck: 6,355, 1 to 14,144
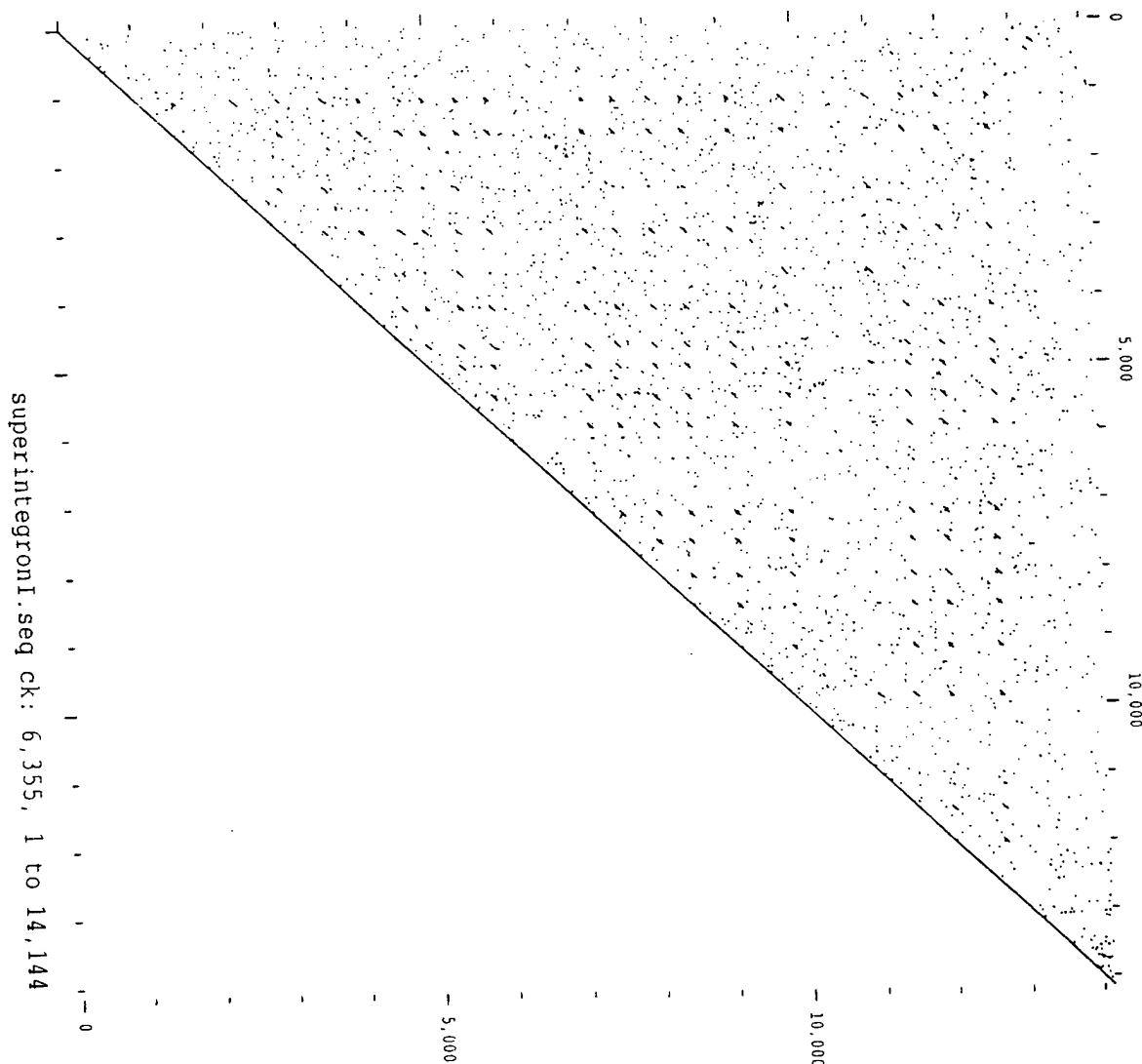
Fig. 4

Fig 5

Squiggle plot of: Par1a.mfold May 26, 1999 11:38

(Linear) MFOLD of: Par1a T: 37.0 Check: 3607 from: 1 to: 78 May 26, 1999 11:38
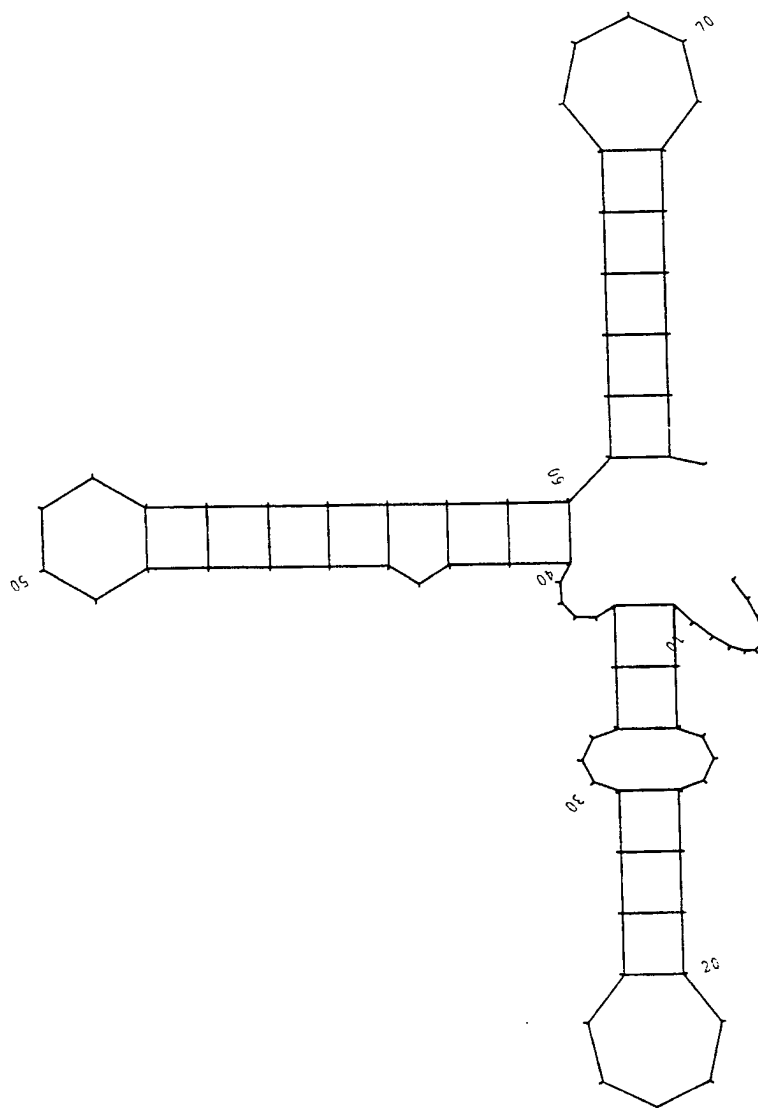
Length: 78 Energy: -16.4

n9/701 626

Fig. 6A

Family 1 of *Pseudomonas alcaligenes* repeat
(PAR) elements

```
Identity    ...TAAC........T.A..........CT.C.TC.CT........C.G..A.GCC......GC.CC..T..C....CGTTA...        (SEQ ID NO:117)
90% (24/27( ..CTAAC.A.TGGTTCAA..........GCTTCGCTCACT.--...GGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGG.      (SEQ ID NO:118)
Majority    GCCTAACAATTGGTTCAA-GCCGTTCGCTCGCTTCGCTCACT---CGGGACCGGCTAAAAGCCG----GCCCCT-TAA-CCAAACGTTAGGC      (SEQ ID NO:119)

PAR1    ATCTAACAATTGGTTCAA-GTCGCTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAGCGTTAGGT      (SEQ ID NO:5)
PAR2    ACCTAACAATGGCGCTCAACCGGCTCCCTTCGGTCGGTCGGTCGGTCGGACCGGCGAGCGCGGTTAGCTCTACGTTAGGC                 (SEQ ID NO:6)
PAR7    ACCTAACAACTGGTTCAA-GTCGTTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGT      (SEQ ID NO:11)
PAR8    GCCTAACAATTGGCTCAAG-TCGTTCGCTTCGCTCACT----CGGGACCGGC--GAAGCCG-----GCCCCT-TAG-CCAAACGTTAGGT      (SEQ ID NO:12)
PAR10   GCCTAACAAGTGGTTCAA--CCGTTCGCTTCGCTCACTG----GGACGGGCTAAAGCCC-----GCCCCT-TAA-CCAAACGTTAGGC        (SEQ ID NO:14)
PAR14   CCCTAACAAATGGTTCAAAAGCCGTTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGAG     (SEQ ID NO:18)
PAR15   CTCTAACAAATGGTTCAA-GTCGCTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGC      (SEQ ID NO:19)
PAR18   CCCTAACTAGTGGTTCAA-GCCGCTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGT      (SEQ ID NO:22)
PAR19   GCCTAACAAATGGTTCAAGTCGTTC-GCTTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGT  (SEQ ID NO:23)
PAR20   ACCTAACAATTGGTTCAA-GTCGTTCGCTTCGCTCACT----CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGT      (SEQ ID NO:24)
PAR22   GCCTAACAATGCGTTCAA-GTCGTTCGCTTCGCTTC-GCTTCGCTCACTCACT---CGGGACCGGCTAAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGT  (SEQ ID NO:26)
PAR34   TTATAACAATTGGTTCAAGTCGTTC-GCTTCGCTCGCTTCGCTCACT---GGGACCGGCTAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGT  (SEQ ID NO:38)
PAR36   GCCTAACAATAGGTTCAA-GTCGCTCGCTTCGCTCACT----CGGGACGGGCTAAAGCCG-----GCCCCT-TAA-CCAAACGTTAGGC       (SEQ ID NO:40)
PAR37   ACCTAACTAGTGGTTCAA-GCCGCTCGCTTCGCTCACT----CGGGACGGGCTAAAGCCG-----GCCCCT-TAA-CCAAACGTTATGC       (SEQ ID NO:41)
PAR38   ATATAACAATTGGTTCAAA-CCGTTCGCTCGCGCTCACTG----GGACGGGCTAAAGCCC-----GCCCCT-TAA-CCAAACGTTATGC       (SEQ ID NO:42)
```

Fig. 6B
Family 2 of *Pseudomonas alcaligenes* repeat
(PAR) elements

```
Identity      ....AAC.....G.TCAA...C.CTC.C...G.TC..T.G.....G.C..........CG...GC.C...TTA.......CGTT...                         (SEQ ID NO:120)
90% (13/14)   ...AACA..GGTTCAA...C.CTC.CT.CGCTC..T.GGGACCG-.C--.......CG--GC.CC--TTA.CCAAACGTT.G.                            (SEQ ID NO:121)
Majority      GCCTAACAACTGGTTCAAG-TCGCTCGCTTCGCTCGCTCGCTCGGGGACCG-GC--TAAAGCCG--TTAACCAAACGTTAGGG                           (SEQ ID NO:122)

PAR2    ACGTAACACATGGCGCTCAACCGGCCGCTCCCTTCGGTCGCTGGACGCGCTGCGCGGATAAAAGCCGGCGCAGCGCCGGTTAGCTCTCTACGTTAGGC  (SEQ ID NO:6)
PAR17   GCCCCAACAAATGGTTCAAG-TCGCTCGCTCCGCTCACTCGGGACCG-GC--TAAAGCCG-----GCCCCC---TTAACCAAACGTTAGGG           (SEQ ID NO:21)
PAR21   GCCTAACAACTGGTTCAAG-TCACTCGGTTCAAG-TCACTCGCTTCGCTCGGGACCG-GC--ATAGC-CG-----GCCCCC---TTAACCAAACGTTAGGT  (SEQ ID NO:25)
PAR28   GCCTAACAACTGGTTCAAG-CCACTCGCTTCGCTCGGGACCG--C--GTACCGCG-----GCCCCC---TTAACCAAACGTTGGGC               (SEQ ID NO:32)
PAR29   GCCCAACAAACGGTTCAAGAACCGCTCGCCTTGCTCGCTCGGGACCG-GC--TAAAACCG-----GCCCCC---TTAACCAAACGTTAGGG          (SEQ ID NO:33)
PAR30   GCCTAACAACTGGTTCAA-ATCGCTCGCTCGCTCGCT-GGGACCG-GC----GAAGCCG-----GCCCCC---TTAACCAAACGTTAGGC           (SEQ ID NO:34)
PAR44   CCCTAACAAATGGTTCAAG-TCACTCGCTTCGCTCGTTCGGGACCG-GC--TAAAGCCG-----GCCCCC---TTAACCAAACGTTAGAG           (SEQ ID NO:48)
PAR53   GCCTAACAACTGGTTCAAG-CCACTCACTTCGCTCGCTCGGGACCG---C--GTTCCGCG-----GCCCCC---TTAACCAAACGTTGGGC          (SEQ ID NO:57)
PAR56   GCCTAACTACTGGTTCAAG-TCACTCGCTTCGCTCGTTCGGGACCG---C--GTTCCGCG-----GCCCCC---TTAACCAAACGTTAGAG          (SEQ ID NO:60)
PAR61   ATCTAACATGTGGTTCAAG-CCGCTCGCTTCGCTCGCTCACTCGGGACCG-GC--TAAAGCCG-----GCCCCC---TTAACCAAACGTTAGAG       (SEQ ID NO:65)
PAR62   CTCTAACAATTGGTTCAAG-CCGCTCGCTCCGCTCGCT-GGGATCG-GC----GAAGCCG------GCACCG---TTAACCAAGCGTTAGAT         (SEQ ID NO:66)
PAR67   GCCTAACAACTGGTTCAA-ATCGCTCGCTCCGCTCGGGGACCG-GC--ATAGCCG------GCCC----TTAGGCAAACGTTAACT              (SEQ ID NO:71)
PAR69   GTTTAACAACTGGTTCAAG-CCGCTCGCTTCGCTCACTCGGGACCG-GC--TAAATTCG-----GCCCCC---TTAGGCAAACGTTAACT           (SEQ ID NO:73)
PAR71   GTATAACAATTGGTTCAAG-TCACTCGCTTCGCTCGCTCGGGACCG-GC--TAAAGCCG-----GCCCCC---TTAACCAAACGTTAGAT           (SEQ ID NO:75)
```

Fig. 6C

Family 3 of *Pseudomonas alcaligenes* repeat (PAR) elements


```
Identity
90% (13/15)
Majority

                10        20        30        40        50        60        70        80

         ...AACA..CGCT AAC..CGCTC.CT.CG.TCGCTGGA........GC.........G.C.GTTAGCT..A.CGTTA...            (SEQ ID NO:1)
         ..CTAACAATGCGCTCAACT..CGCTCACTTCGTTCGCTGGAC...CAAAAAGCT.CGCTTTTG..GCCCGTTAGCTTAATCGTTA...    (SEQ ID NO:1)
         ACCTAACAATGCGCTCAACTGTCGCTGGCTGGACAGCCAAAAAGCTGCGCTTTTGCCTGCCCGTTAGCTTAATCGTTAGGC           (SEQ ID NO:125)

PAR5    ACCTAACAATGCGCTCAACTGCCGCTCACTTCGTTCGCTGGACAGTCGGCTGGCTCAAAAAGCTGCGCTTTTGCCTGCCCGTTAGCTTAATCGTTAGGC    (SEQ ID NO:9)
PAR9    ACCTAACAATGCGCTCAACTGCCGCTCACTTCGTTCGCTGGACAGTCAAAAAGCTCAAAAAGCTGCGCTTTTGCCTGCCCGTTAGCTTAATCGTTAGGC    (SEQ ID NO:13)
PAR32   ACCTAACAATGCGCTCAACTGTCGCTCGCTCACTTCGTTCGCTGGACAGTCAAAAAGCTCAAAAAGCTGCGCTTTTGCCTGCCCGTTAGCTTAATCGTTAGCG    (SEQ ID NO:36)
PAR41   GCCTAACAATGCGCTCAACTGCCGCTCACTTCGTTCGCTGGACAGTCAAAAAGCTCAAAAAGCTGCGCTTTTGCCTGCCCGTTAGCTTAATCGTTAGAG    (SEQ ID NO:45)
PAR25   ACATAACAATGCGCTCAACTGTCGCTCACTTCGTTCGCTGGACAGCCAAAAAGTACGCTTTTGCTGCCCGTTAGCTTAATCGTTAGGC    (SEQ ID NO:29)
PAR13   ATTTAACAATGCGCTCAACTGTCGCTCACTTCGTTCGCTGGACAGCCAAAAAGCTGCGCTTTTGTCTGCCCGTTAGCTTAATCGTTAGGG    (SEQ ID NO:17)
PAR52   GGCTAACAATGCGCTCAACTGTCGCTCACTTCGTTCGCTGGACAGTCAAAAAGCTACGCTTTTGTCTGCCCGTTAGCTTAATCGTTAGGA    (SEQ ID NO:56)
PAR66   TGCTAACAATGCGCTTAACTGTCGCTCACTTCGTTCGCTGGATAGTCAAAAAGCTGCGCTTTTGTCTGCCCGTTAGCTTAATCGTTAGGC    (SEQ ID NO:70)
PAR26   GCCTAACAAGTCGCTCAACTGTCGCTCACTCCGTTCGCTGGACAGCCAAAAAGCTGCGCTTTTGTCTGCCCGTTAGCTTAATCGTTAGGC    (SEQ ID NO:30)
PAR24   GCCTAACAATACGCTCAACTATCGCTCACTATCGTTCGCTGGACGTCCAAAAAGCTGCGCTTTTGGCCGCCCGTTAGCTTAACCGTTATGT    (SEQ ID NO:31)
PAR27   GCCTAACAATGCGCTCAACTGCCGCTCACTATCGTTCGCTGGACGTCCAAAAAGCTACGCTTTTGGCCGCCCGTTAGCTTAATCGTTATGC    (SEQ ID NO:58)
PAR54   GCCCAACAATGCGCTCAACTGCCGCTCACTTCGTTCGCTGGACGTCCAAAAAGCTCGCAAAAAGCTGCGCTTTTGCTCGCCCGTTAGCTTAATCGTTAGGT    (SEQ ID NO:8)
PAR4    ACCTAACAATGCGCTCAACTGCCGCTCACTTCGTTCGCTGGACTCGCAAAAAGCTGCGCTTTTGCTCGCCCGTTAGCTTAATCGTTATAA    (SEQ ID NO:15)
PAR11   CTCTAACAATGCGCTCAACTATCGCTCACTTCGTTCGCTGGACTCGCAAAAAGCTGCGCTTTTGCTCGCCCGATAAAAGCCGCGCGATAAAAGCCGCGCCAGCGCGTTAGCTCTA-CGTTAGGC    (SEQ ID NO:6)
PAR2    ACCTAACAATGGCGCTCAACCG--CGCTCCCCTTCGGTCGCTGGAACGCTGCGCGATAAAAGCCGCGCCAGCCGGTTAGCTCTA-CGTTAGGC    (SEQ ID NO:6)
```

Fig. 6D
Family 4 of *Pseudomonas alcaligenes* repeat
(PAR) elements


```
                          10        20        30        40        50        60        70        80
Identity     ...TAACA...CGCTCAA...CG.TC.CT.CG.T.C.CTGG....G.GC....AAGCC......GC.CC..TTAGCT...CGTTA...     (SEQ ID NO:126)
90% (7/8)    ...TAACAA.GCGCTCAAA-.CGCTC.CTTCG.T-CGCTGGG-AC.G-GC---.AAGCC.----GCCCC--TTAGCTTAATCGTTAG..     (SEQ ID NO:127)
Majority     GCCTAACAATGCGCTCAAA-GCGCTCACTTCGTT-CGCTGGG-ACCG-GC--TAAAGCCG----GCCCC--TTAGCTTAATCGTTAGGT     (SEQ ID NO:128)

PAR6         GCCTAACAATGCGCTCAAA-GCGCTCACTTCGTT-CGCTGGG-ACCG-GC----GAAGCCG----GCCCC---TTAGCTTAATCGTTAGGT   (SEQ ID NO:10)
PAR31        GCCTAACAATGCGCTCAAA-GCGCTCACTTCGTT-CGCTGGG-ACCG-GC----GAAGCCG----GCCCC---TTAGCTTAATCGTTAGGT   (SEQ ID NO:35)
PAR50        GCCTAACAATGCGCTCAAA-GCGCTCACTTCGTT-CGCTGGG-ACCG-GC----GAAGCCG----GCCCC---TTAGCTTAATCGTTAGAA   (SEQ ID NO:54)
PAR65        GTCTAACAATGCGCTCAAA-GCGCTCACTTCGTT-CGCTGGG-ATCG-GC----TAAAGCCG----GCCCC---TTAGCTTAATCGTTAGCA   (SEQ ID NO:69)
PAR42        CTCTAACAATGCGCTCAAA-TCGCTCACTACGTT-CGCTGGG-ACCG-GC----TAAAGCCG----GCCCC---TTAGCTTAATCGTTAGAG   (SEQ ID NO:46)
PAR72        GCATAACAAGTCGCTCAAA-TCGCTCACTTCGTTCGCTGGG-ACGG-GC----TAAAGCCC----GCCCC---TTAGCTTA-TCGTTAGGC   (SEQ ID NO:76)
PAR12        TTATAACAATGCGCTCAAA-TCGTTCGCTTCGCT-CACTGGG-ACGG-GC----TAAAGCCC----GCCCC---TTAGCTTAATCGTTAAAT   (SEQ ID NO:16)
PAR2         ACCTAACATGGCGCTCAACCGCGCTCCCTTCGGT-CGCTGGACGCTGCGCGATAAAGCCGGCAGCGCCCGGTTAGCTCTA-CGTTAGGC   (SEQ ID NO:6)
```

Fig 7A. PAR-specific oligonucleotide (bottom) aligned with PAR majority consensus (top)

```
majority  GCCTAACAATTGGTTCAAG-GTCGCTCGCTTCGCTCACT-CGGGACCGGCTAAAGC-------CGGCC--CC-TTAA-CCAAACGTTAGGC (SEQ ID NO:129)

oligo 1                     5' TCGCTTCGCTCACTGCGGGACCGGCTCACTGCGGGACCG                    -------CC--TTAA-CCAAACGTTA   (SEQ ID NO:79)
       2      5' TAACAATTGGTTCAAG--TCGTTCGCTCACTGCGGGACCG                                               (SEQ ID NO:80)
       3      5' TAACTATTCAGTCAAGCGGA---CGCAAACCCGCTGCGCGGTCTT                                          (SEQ ID NO:81)
       4      5' TAACAATGCGCTCAACTGCG-CTCACTTCGTTCGCTGGACAGCC                                           (SEQ ID NO:82)
       5      5' TAACAAGTCGCTCAACTGCCGCTCACT-CGTTCGCTGGACAGCC                                           (SEQ ID NO:83)
```

Fig. 7B. PCR primers for PAR fingerprints

```
PAR majority:         5' GCCTAACAATTGGTTCAAG-GTCGCTCGCTTCGCTCACT-CGGGACCGGCTAAAGC                        (SEQ ID NO:129)

Oligo 6 (complement of PAR):                  3' GCGAGCGAAGCGAGTGA-GCC                                     (SEQ ID NO:84)

PAR majority (cont'd):    -------CGGCC--CC-TTAA-CCAAACGTTAGGC                                     Continuation of
                                                                                                (SEQ ID NO:129)
Oligo 7:                       GCC--CC-TTAA-CCAAACGTTAG 3'                                       (SEQ ID NO:85)
```
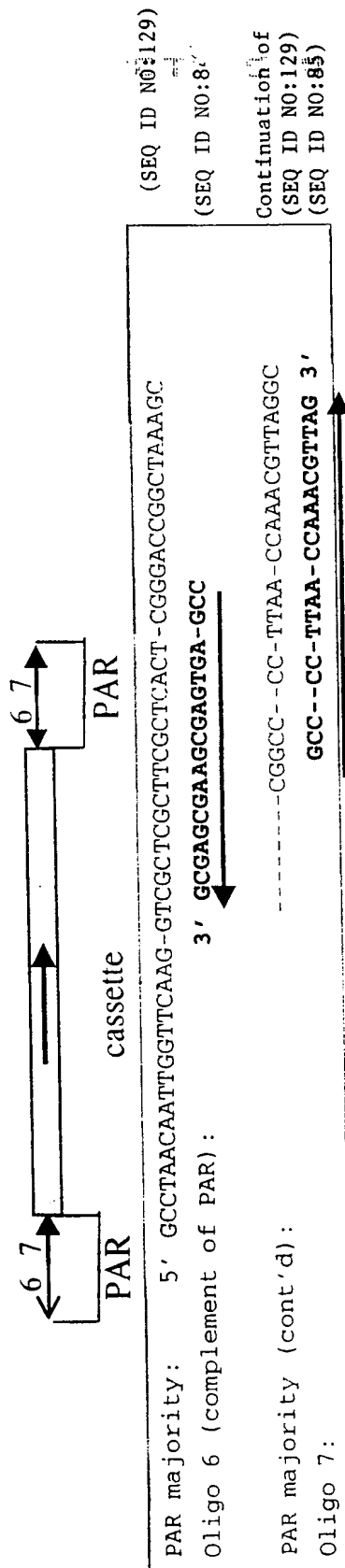
Fig 8. Hybridization of a PAR-specific Oligonucleotide 1 to
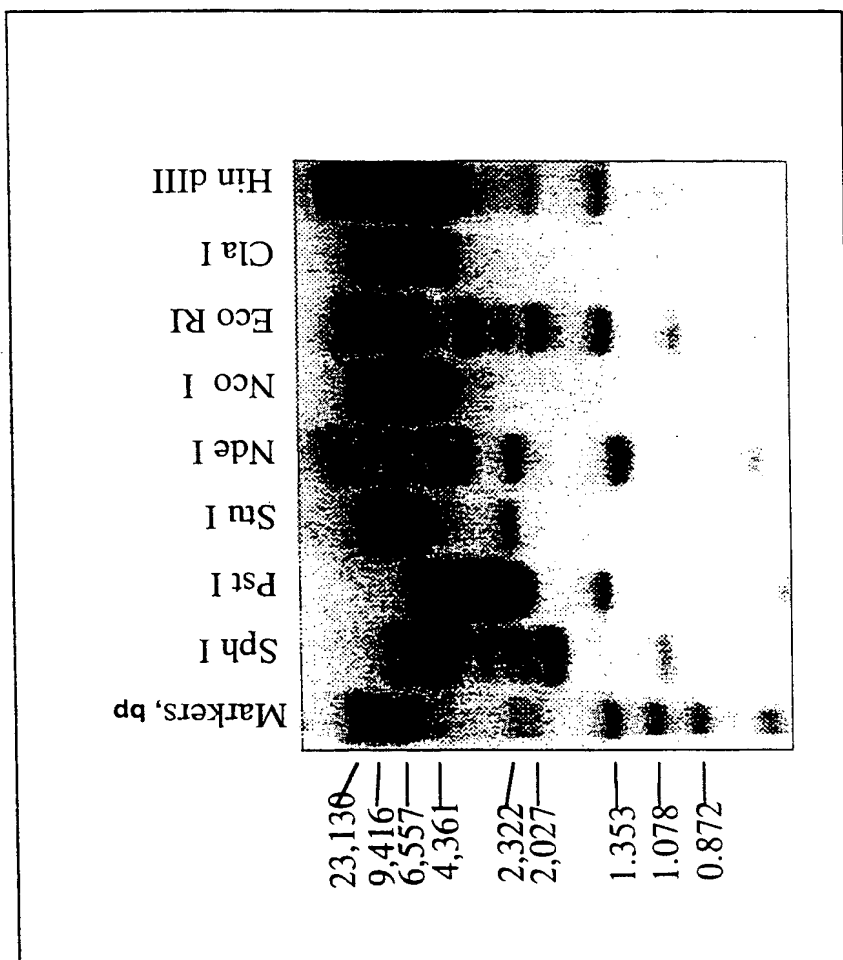*Pseudomonas alcaligenes* chromosomal DNA.

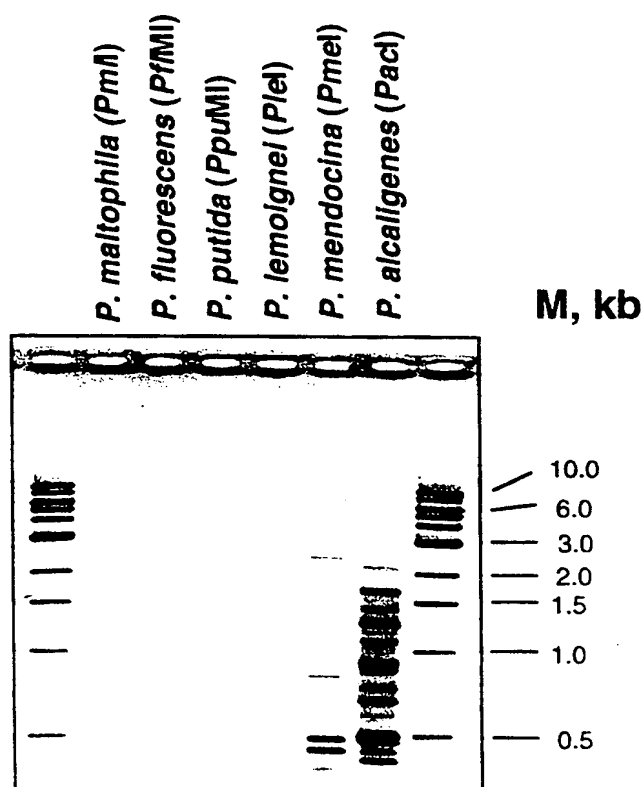# Fig.9. Distribution of PAR Cassettes Among *Pseudomonas* Species

09/701626

Fig. . Amplification and cloning strategy